

TRABALHO DE GRADUAÇÃO

Método para Detecção Automática de Nudez em Imagens  
Digitais utilizando Redes Neurais Artificiais

Daniel Assad Maia Sandoval

Brasília, Julho de 2014

UNIVERSIDADE DE BRASÍLIA

INSTITUTO DE CIÊNCIAS EXATAS  
FACULDADE DE TECNOLOGIA

Universidade de Brasília — UnB  
Instituto de Ciências Exatas / Faculdade de Tecnologia  
Departamento de Ciência da Computação / Departamento de Engenharia Elétrica  
Bacharelado em Engenharia de Computação

Banca examinadora composta por:

Prof. Dr. Alexandre Zaghetto (Orientador) — CIC/UnB  
Prof. Dr. Bruno Macchiavello — CIC/UnB  
Prof. Dr. Marcus V. Chaffim Costa — FGA/UnB

### **CIP — Catalogação Internacional na Publicação**

Sandoval, Daniel Assad Maia.

Método para Detecção Automática de Nudez em Imagens Digitais utilizando Redes Neurais Artificiais / Daniel Assad Maia Sandoval. Brasília : UnB, 2014.

61p. : il. ; 297mm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2014.

1. Processamento Digital de Imagens, 2. Detecção de Nudez, 3. Redes Neurais Artificiais, 4. Classificação de Padrões, 5. Segmentação por Cor de Pele

CDU 621.3

Endereço: Universidade de Brasília  
Campus Universitário Darcy Ribeiro — Asa Norte  
CEP 70910-900  
Brasília-DF — Brasil

UNIVERSIDADE DE BRASÍLIA  
Faculdade de Tecnologia e Instituto de Ciências Exatas

TRABALHO DE GRADUAÇÃO

**Método para Detecção Automática de Nudez em Imagens  
Digitais utilizando Redes Neurais Artificiais**

**Daniel Assad Maia Sandoval**

*Relatório submetido ao Departamento de Ciência da  
Computação e ao Departamento de Engenharia  
Elétrica como requisito parcial para obtenção  
do grau de Engenheiro de Computação*

Banca Examinadora

Prof. Dr. Alexandre Zaghetto, CIC/UnB  
*Orientador*

\_\_\_\_\_

Prof. Dr. Bruno Macchiavello, CIC/UnB  
*Examinador interno*

\_\_\_\_\_

Prof. Dr. Marcus V. Chaffim Costa, FGA/UnB  
*Examinador externo*

\_\_\_\_\_

*Dedico este trabalho à minha família. Em especial, à minha mãe, ao meu pai, à minha irmã e aos meus avós.*

*Daniel Assad Maia Sandoval*

## Agradecimentos

*Agradeço à minha mãe, **Anna Beatriz Assad Maia**, e ao meu pai, **Wilson Giometti Sandoval**, por me guiarem até aqui com a dedicação e carinho que tornaram este momento possível.*

*Agradeço à minha irmã, **Letícia Assad Maia Sandoval**, pelas conversas, piadas e lutas que tornaram este processo mais divertido.*

*Agradeço aos meus avós, **Marízia Assad Alvez Maia**, **Roberto Alves Maia**, **Helena Levy Giometti Sandoval** e **Wilson Porto Sandoval** por sua luta e garra, que por fim subsidiaram meus estudos. Pelos ensinamentos e contos transmitidos com tanto amor.*

*Agradeço à **Zélia Júlia de Jesus** por escolher se juntar à minha família e trazer, entre outras alegrias, a **Victória de Jesus Ribeiro**.*

*Aos demais familiares, tão importantes quanto os aqui nomeados, entre madrinhas e padrinhos, tias e tios, primas e primos, agradeço pelo ambiente criado, pelas lições e brincadeiras.*

*Agradeço ao Professor **Alexandre Zaghetto** por ter me acolhido como seu aluno desde cedo, guiando-me sempre de maneira leve a fazer o correto. Também deixo meus agradecimentos aos demais Professores que tão importantes foram neste processo.*

*Agradeço aos meus amigos pela companhia na jornada.*

*Daniel Assad Maia Sandoval*

---

## RESUMO

O presente trabalho apresenta solução baseada em redes neurais artificiais para classificação de imagens digitais em dois grupos: (i) imagens que contém nudez; e (ii) imagens que não contém nudez. A solução apresentada pode ser dividida em dois módulos. O primeiro classifica *pixels* da imagem em pele ou não-pele. O segundo extrai características e classifica entre nudez ou não. Ambos utilizam redes neurais artificiais. O método proposto obteve *F measure* de 0.8735 utilizando o espaço de cor *RGB* para classificação de *pixels* de pele. Também é apresentada revisão bibliográfica de técnicas utilizadas e trabalhos semelhantes.

---

## ABSTRACT

Neural-network based solution for digital image classification in one of two groups: (i) images that contain nudity; and (ii) images that do not contain nudity. The proposed solution is composed of two modules. The first module classifies the image's pixels as skin or non-skin pixels. The second extracts characteristics and classifies images as containing nudity or not containing nudity. Both utilize artificial neural networks. The propose method achieved F measure of 0.8735 using the *RGB* color space for skin segmentation. It's also presented a revision of utilized techniques and similar works.

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>1</b>
1.1	CONTEXTUALIZAÇÃO .....	1
1.2	DEFINIÇÃO DO PROBLEMA .....	3
1.3	OBJETIVOS DO PROJETO.....	4
1.4	FERRAMENTAS DE IMPLEMENTAÇÃO E PSEUDOCÓDIGO APRESENTADO .....	4
1.5	APRESENTAÇÃO DO MANUSCRITO .....	4
<b>2</b>	<b>PROCESSAMENTO DIGITAL DE IMAGENS.....</b>	<b>6</b>
2.1	IMAGEM DIGITAL .....	6
2.2	OPERAÇÕES MORFOLÓGICAS .....	7
2.2.1	DILATAÇÃO.....	7
2.2.2	EROSÃO.....	8
2.2.3	ABERTURA E FECHAMENTO.....	8
2.3	SEGMENTAÇÃO DE IMAGENS .....	9
2.3.1	SEGMENTAÇÃO POR COR DE PELE .....	10
2.4	ESPAÇOS DE COR.....	11
2.4.1	<i>RGB</i> .....	11
2.4.2	<i>nRGB</i> .....	12
2.4.3	<i>nRG</i> .....	13
2.4.4	<i>HSV</i> .....	13
2.4.5	<i>YC<sub>b</sub>C<sub>r</sub></i> .....	14
<b>3</b>	<b>RECONHECIMENTO DE PADRÕES .....</b>	<b>16</b>
3.1	ASPECTOS GERAIS DE SISTEMAS DE RECONHECIMENTO DE PADRÕES.....	17
3.1.1	MODELO.....	17
3.1.2	PRÉ-PROCESSAMENTO E SEGMENTAÇÃO .....	17
3.1.3	EXTRAÇÃO DE CARACTERÍSTICAS E AMOSTRAS DE TREINAMENTO .....	17
3.1.4	CUSTO DO ERRO E TEORIA DE DECISÃO .....	18
3.2	FUNÇÕES DISCRIMINANTES LINEARES .....	18
3.2.1	FUNÇÕES DISCRIMINANTES LINEARES PARA CLASSIFICADORES COM DUAS CA- TEGORIAS .....	19
3.3	REDES NEURAIS ARTIFICIAIS .....	20
3.3.1	NEURÔNIO.....	20

3.3.2	REDES NEURAIS ARTIFICIAIS DE MÚLTIPLAS CAMADAS .....	23
3.4	DETECÇÃO DE FACES.....	25
3.5	RECONHECIMENTO DE PADRÕES MORFOLÓGICOS.....	26
3.6	MÉTODO PARA AVALIAÇÃO DOS CLASSIFICADORES UTILIZADOS.....	27
3.6.1	<i>Precision</i> E <i>recall</i> .....	27
3.6.2	<i>F measure</i> .....	28
<b>4</b>	<b>ESTUDO DA DISTRIBUIÇÃO DE CORES DE PELE.....</b>	<b>29</b>
4.1	BASE DE TESTES .....	29
4.2	DISTRIBUIÇÃO DOS GRUPOS NOS ESPAÇOS DE COR ESTUDADOS .....	30
4.2.1	DISTRIBUIÇÃO NO ESPAÇO DE COR <i>RGB</i> .....	30
4.2.2	DISTRIBUIÇÃO NOS ESPAÇOS DE COR <i>nRGB</i> E <i>nRG</i> .....	30
4.2.3	DISTRIBUIÇÃO NO ESPAÇO DE COR <i>HS</i> .....	33
4.2.4	DISTRIBUIÇÃO NO ESPAÇO DE COR <i>C<sub>b</sub>C<sub>r</sub></i> .....	33
4.3	SEGMENTAÇÃO DE COR DE PELE POR FUNÇÕES DISCRIMINANTES LINEARES ...	34
4.3.1	CLASSIFICADOR PROPOSTO.....	34
4.3.2	RESULTADOS OBTIDOS .....	35
4.4	SEGMENTAÇÃO DE COR DE PELE COM REDES NEURAIS ARTIFICIAIS .....	35
4.4.1	CLASSIFICADOR PROPOSTO.....	35
4.4.2	RESULTADOS OBTIDOS .....	36
<b>5</b>	<b>MÉTODO PROPOSTO.....</b>	<b>38</b>
5.1	SOLUÇÃO PROPOSTA ANTERIORMENTE.....	38
5.2	CLASSIFICADOR DE COR DE PELE.....	39
5.3	EXTRATOR DE CARACTERÍSTICAS.....	42
5.3.1	NÚMERO DE ELEMENTOS CONECTADOS DE ÁREA SIGNIFICATIVA .....	42
5.3.2	ÁREA RELATIVA DOS DOIS MAIORES ELEMENTOS CONECTADOS.....	43
5.3.3	ÁREA RELATIVA DE PELE EM ZONAS .....	43
5.3.4	CARACTERÍSTICAS DE FACES PRESENTES NA IMAGEM.....	43
5.3.5	FORMAS DE ELEMENTOS CONECTADOS DE PELE.....	45
5.4	CLASSIFICADOR DE NUDEZ .....	46
<b>6</b>	<b>RESULTADOS OBTIDOS.....</b>	<b>49</b>
6.1	CONJUNTO DE TESTES .....	49
6.2	PROTOCOLO DE TESTES .....	49
6.3	FASE <i>i</i> – AVALIAÇÃO DE CARACTERÍSTICAS .....	50
6.3.1	ÁREA MÍNIMA PARA CONTAGEM DE ELEMENTOS CONECTADOS .....	50
6.3.2	ÁREA RELATIVA DOS MAIORES ELEMENTOS CONECTADOS .....	50
6.3.3	ÁREA RELATIVA DE PELE EM ZONAS DA IMAGEM .....	51
6.3.4	CARACTERÍSTICAS DE FACES PRESENTES NA IMAGEM.....	51
6.3.5	CARACTERÍSTICAS DA FORMA DOS ELEMENTOS CONECTADOS .....	51
6.3.6	CONCLUSÃO DA FASE <i>i</i> DE TESTES.....	53
6.4	FASE <i>ii</i> – AVALIAÇÃO DO SISTEMA PROPOSTO .....	53



<b>7</b>	<b>CONCLUSÕES .....</b>	<b>55</b>
7.1	OBJETIVO GERAL .....	55
7.2	OUTRAS CONTRIBUIÇÕES .....	55
7.3	TRABALHOS FUTUROS.....	56
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>57</b>
	<b>ANEXOS.....</b>	<b>60</b>
<b>I</b>	<b>DESCRIÇÃO DO CONTEÚDO DO CD .....</b>	<b>61</b>

# LISTA DE FIGURAS

2.1	Representação digital de uma imagem. ....	7
2.2	Dilatação e erosão de um quadrado azul escuro por um círculo. ....	8
2.3	Abertura e fechamento de formas em azul escuro por um círculo. ....	9
2.4	Segmentação de corpos d'água em uma imagem de satélite de Brasília-DF. ....	10
2.5	Representação do espaço de cores RGB. ....	12
2.6	Representação do espaço de cores HSV. ....	13
2.7	Distribuição de cores no espaço $C_bC_r$ para diferentes valores de $Y$ . ....	15
3.1	Distribuição de indivíduos de dois grupos distintos no espaço de suas características. ....	19
3.2	Representação de neurônio biológico. ....	21
3.3	Representação de neurônio MCP. ....	21
3.4	Funções de ativação degrau e signum. ....	22
3.5	Função de ativação sigmóide com diferentes valores de $\alpha$ . ....	23
3.6	Função de ativação tangente hiperbólica. ....	23
3.7	Rede neural completamente ligada com dois neurônios de entrada, cinco neurônios na camada escondida e um neurônio de saída. ....	24
3.8	Processo de afinamento de imagens binárias. ....	26
4.1	Distribuição de <i>pixels</i> de pele e não pele no espaço de cores $RGB$ . ....	31
4.2	Distribuição de <i>pixels</i> de pele e não pele no espaço de cores $nRGB$ . ....	32
4.3	Distribuição de <i>pixels</i> de pele e não pele no espaço de cores $HS$ . ....	33
4.4	Distribuição de <i>pixels</i> de pele e não pele no espaço de cores $C_bC_r$ . ....	34
5.1	Diagrama ilustrativo do método proposto. ....	38
5.2	Exemplo da segmentação por cor de pele. ....	41
5.3	Exemplo de identificação de elementos conectados. ....	41
5.4	Definição de zonas da imagem. ....	44
5.5	Exemplo de extração da característica de <i>pixels</i> de pele presentes em faces. ....	47
5.6	Exemplificação do processo de extração de características morfológicas. ....	47
5.7	Método proposto completo com detalhamento das características utilizadas. ....	48

# LISTA DE TABELAS

3.1	Classificação de elementos para avaliação de eficácia. ....	28
4.1	Tabela representativa dos resultados obtidos com segmentação de pele por menor distância ao centróide. ....	35
4.2	Análise de eficácia das RNAs de classificação de pele para diferentes espaços de cor ..	36
6.1	Análise de eficácia com diferentes limiares $\rho$ para área significativa. ....	52
6.2	Análise de eficácia com área dos $n$ maiores elementos conectados. ....	52
6.3	Análise de eficácia com área de pele por em zonas. ....	52
6.4	Análise de eficácia com características de faces na imagem. ....	52
6.5	Análise de eficácia com informações da forma dos elementos conectados. ....	54
6.6	Eficácia da solução proposta com diferentes espaços de cor para classificação de nudez.	54

# LISTA DE SÍMBOLOS

## Siglas

OCR	Reconhecimento Ótico de Caracteres (do inglês, <i>Optical Character Recognition</i> )
RNA	Rede Neural Artificial
MCP	Neurônio de McCulloch-Pitts
MP	Milhões de <i>pixels</i> ( <i>megapixels</i> )
DSLR	Câmera do tipo <i>digital single-lens reflex</i>
CIPA	<i>Camera &amp; Imaging Products Association</i>
JPEG	<i>Joint Photographic Experts Group</i>
MPEG	<i>Moving Picture Experts Group</i>
ISO	<i>International Organization for Standardization</i>
IEC	<i>International Electrotechnical Commission</i>
CRT	<i>Cathode Ray Tube</i>
LCD	<i>Liquid-Crystal Display</i>
LED	<i>Light-Emitting Diode</i>
DLP	<i>Digital Light Processing</i>
LCOS	<i>Liquid Crystal on Silicon</i>
$RGB$	Espaço de cor RGB, <i>Red Green Blue</i>
$nRGB$	<i>Normalized RGB</i> , ou espaço de cor RGB normalizado
$nRG$	Espaço de cor dos componentes $R$ e $G$ do espaço $nRGB$
$YC_bC_r$	Espaço de cor $YC_bC_r$
$C_bC_r$	Espaço de cor dos componentes $C_b$ e $C_r$ do espaço $YC_bC_r$
$HSV$	Espaço de cor $HSV$ , <i>Hue Saturation Value</i>
$HS$	Espaço de cor dos componentes <i>Hue</i> e <i>Saturation</i> do espaço $HSV$
$CIELab$	Espaço de cor $CIELab$ , também conhecido como <i>Lab</i>

# Capítulo 1

## Introdução

*“Il semble que la perfection soit atteinte non quand il n’y a plus rien à ajouter, mais quand il n’y a plus rien à retrancher.” —Antoine de Saint Exupéry*

### 1.1 Contextualização

O ser humano sempre buscou meios de registrar sua história. Já com as pinturas rupestres, que datam de dezenas de milhares de anos no passado [1], buscamos, inclusive, maneiras gráficas de representar desde eventos do cotidiano a grandes feitos da humanidade. À medida que construímos o conhecimento científico, buscamos criar dispositivos que nos auxiliassem na tarefa de registrar o que vemos a fim de recordar ou reproduzir a mesma experiência sensorial de visão posteriormente.

Em 1544 o matemático Reiniers Gemma Frisius utilizou uma *camara obscura*, precursora da câmera fotográfica, para observar um eclipse solar e descreveu seu método em uma publicação científica. Em 1558, Giovanni Batista della Porta sugeriu que este método fosse utilizado como um auxílio para desenhar, como uma forma de transcrever o que era visto [2]. Desde então fizemos grande e rápido avanço no projeto e desenvolvimento de dispositivos capazes de registrar imagens.

Dos principais avanços na área de fotografia podemos destacar a utilização pioneira de filme fotográfico por George Eastman, que colocou à venda sua primeira câmera em 1888, a Kodak<sup>1</sup>. Esta inovação permitiu a expansão das câmeras fotográficas ao mercado consumidor, a ponto de se tornarem itens comuns nos anos 1900. A criação deste mercado possibilitou vasta pesquisa e desenvolvimento na área e muito se evoluiu na fotografia com base na popularização de dispositivos com filme fotográfico.

No entanto, podemos destacar outro evento divisor de águas na fotografia: a invenção da máquina fotográfica digital. Em 1975 Steve Sasson, então engenheiro na Kodak, desenvolveu o primeiro protótipo de máquina digital. Este pesava  $3.6kg$  e demorou 23 segundos para tirar uma foto preta e branca com  $0.01MP$  (dez mil *pixels*) e armazená-la em uma fita cassete. O caminho ainda era longo para uma versão comercial de uma câmera digital, até que em 1990 surgiu a *Dycam Model 1*, a primeira câmera digital disponível no mercado. Esta armazenava imagens digitalmente

---

<sup>1</sup>Informação disponível em [kodak.com](http://kodak.com)

e se conectava a um computador para transferí-las.

Desde 1990, em apenas 24 anos, as câmeras digitais passaram de hobby de poucos entusiastas que dispunham do interesse necessário para obter uma foto digital com resolução ínfima e 256 níveis de cinza para um artefato presente no cotidiano de todos, seja apenas como câmeras ou incorporadas nos mais diversos produtos, como computadores pessoais e telefones celulares. Este fenômeno foi possível devido ao grande avanço tecnológico da produção de circuitos eletrônicos, o que acabou por reduzir os custos de produção e aumentar a capacidade de processamento e armazenamento de dados.

Dois grandes avanços tecnológicos em outras áreas também permitiram a difusão das câmeras digitais, como a criação do padrão de compressão de imagens *JPEG* [3] e a ampla difusão da internet. Aquele peritiu que mais imagens fossem armazenadas na memória, à época muito mais escassa e cara, enquanto este aumentou a demanda por imagens digitais, uma vez que era a única forma de transmitir fotografias eletronicamente.

Hoje em dia câmeras digitais tornaram-se itens comuns a grande parcela da população, a ponto em que as encontramos em 83% dos 5,2 bilhões de telefones celulares em uso no mundo<sup>2</sup>. Além disso, câmeras digitais profissionais ou semiprofissionais, como as DSLRs (câmeras *digital single-lens reflex*), capazes de adquirir fotos de alta resolução e qualidade, tiveram mais de 17 milhões de unidades vendidas em 2013<sup>3</sup>.

Também nos últimos anos testemunhamos a rápida ascensão do uso da internet por cada vez mais pessoas. A internet tornou-se um meio de comunicação vastamente utilizado para qualquer interação humana, seja esta social, por meio de aplicações de mensagens instantâneas ou redes sociais, ou comercial, como a de um cliente que compra de uma empresa pela internet. Fotografias digitais tem um importante papel nas interações que ocorrem na internet, sendo inclusive um forte atrativo, por conta do estímulo visual provido. O Facebook, um popular serviço de rede social, já armazena mais de 250 bilhões de fotos de seus usuários, recebendo cerca de 350 milhões de novas imagens todos os dias [4].

Neste cenário, qualquer indivíduo é capaz de adquirir fotos digitais com qualidade profissional e armazená-las em grandes quantidades. Ainda mais, qualquer indivíduo é também capaz de compartilhá-las facilmente por meio da internet. Temos, portanto, dois contextos muito importantes para situar este trabalho em suas contribuições: *(i)* grandes coleções particulares de fotografias digitais; *(ii)* grande volume de imagens trafegando em serviços da internet.

## Pornografia

A palavra pornografia vem do grego *pornographos*, que significa, originalmente, escritos sobre prostitutas; referência à vida, costumes e hábitos das prostitutas e de seus clientes. Segundo o dicionário Houaiss [5], pornografia é definida como “estudo da prostituição; coleção de pinturas ou gravuras obscenas; característica do que fere o pudor (numa publicação, num filme etc.); obsceni-

---

<sup>2</sup>Dados de 2012 do *2013 TomiAhonen Almanac*, disponível em [communities-dominate.blogspot.com](http://communities-dominate.blogspot.com)

<sup>3</sup>Dados da CIPA (*Camera & Imaging Products Association*), disponíveis em [cipa.jp](http://cipa.jp)

dade, indecência, licenciosidade; qualquer coisa feita com o intuito de ser pornográfico, de explorar o sexo tratado de maneira chula, como atrativo (p.ex., revistas, fotografias, filmes etc.); violação ao pudor, ao recato, à reserva, socialmente exigidos em matéria sexual; indecência, libertinagem, imoralidade.”

Justamente por conta de seu caráter obsceno e que fere o pudor, a identificação de imagens pornográficas pode ser útil. Por exemplo, em um ambiente profissional, como uma rede corporativa com acesso à internet, pode ser desejável filtrar conteúdos impróprios. Muitos filtros já são utilizados em servidores *proxy* com esta finalidade, porém geralmente atuam bloqueando endereços previamente indicados como inapropriados. No entanto, há casos em que o mesmo provedor apresenta ambos tipos de conteúdo, próprios e impróprios. Nestes casos filtros simplistas como de endereços não são suficientes e torna-se necessário analisar o conteúdo em si. Neste contexto é útil um sistema automatizado de classificação de imagens pornográficas.

Outro problema relacionado a imagens pornográficas, inclusive mais premente, está relacionado à pornografia infantil. No Estatuto da Criança e do Adolescente<sup>4</sup>, é utilizado o termo “cena de sexo explícito ou pornográfica,” definido como “qualquer situação que envolva criança ou adolescente em atividades sexuais explícitas, reais ou simuladas, ou exibição dos órgãos genitais de uma criança ou adolescente para fins primordialmente sexuais.” Esta é a definição de pornografia infantil nos termos da lei brasileira, que determina ilegal a posse ou distribuição de qualquer material dessa natureza.

A pornografia infantil é um problema não só no Brasil mas também no mundo inteiro, havendo esforços nacionais e internacionais de identificação e acusação judicial dos indivíduos que participam em atividades de pedofilia, entre elas, a posse e distribuição de pornografia infantil. Neste contexto há indivíduos suspeitos de cometer tais crimes que possuem vastas coleções de imagens e vídeos digitais. Ao servir um mandato de busca e apreensão, agentes policiais devem ser capazes de encontrar pornografia infantil no local para que ocorra a prisão em flagrante, caso contrário seria o processo judicial muito prejudicado. Portanto, uma ferramenta automática capaz de identificar imagens de pornografia infantil ou, ao menos, imagens pornográficas, é de grande utilidade aos agentes policiais por diminuir a quantidade de amostras que devem ser manualmente analisadas pelos agentes. Pode-se citar, inclusive, uma ferramenta desenvolvida pela Polícia Federal Brasileira que se propõe a classificar imagens que contém nudez [6].

## 1.2 Definição do problema

Levando-se em consideração o contexto discutido na Seção 1.1, é de grande utilidade uma ferramenta capaz de identificar imagens pornográficas dentre um conjunto de imagens arbitrariamente selecionadas. Tal ferramenta pode ser utilizada tanto em casos extremos como auxiliar agentes policiais em investigações criminais até filtros de conteúdo mais inteligentes, que atuem sobre o próprio conteúdo que será ou não apresentada ao usuário.

Portanto, o problema abordado neste trabalho é: *dado um conjunto de imagens arbitrariamente*

---

<sup>4</sup>Lei nº 8.069/90 e emendas, disponíveis em [planalto.gov.br](http://planalto.gov.br)

*escolhidas, identificar de maneira automática as que apresentem conteúdo pornográfico.*

### 1.3 Objetivos do projeto

Conforme definição fornecida ainda na Seção 1.1, percebe-se que pornografia é um conceito altamente subjetivo, pautado na moral vigente a fim de que se estabeleçam características como obscenidade, ou violação do pudor, para então se classificar como pornográfico. Portanto, como não é possível determinar de maneira objetiva se uma imagem é pornográfica, fazer um classificador automático capaz de tomar esta decisão não é tarefa fácil. A fim de tornar mais objetivo o resultado almejado, foi assumido que imagens pornográficas possuem como principal característica a presença de nudez.

O objetivo principal geral trabalho é, portanto, *propor um sistema capaz de classificar imagens em dois grupos: (i) imagens que contém nudez; e (ii) imagens que não contém nudez.*

Ainda mais, conforme melhor explorado no Capítulo 4, o primeiro passo do método proposto é a segmentação de imagens por cor de pele. Com isto e demais aspectos em mente, foram definidos objetivos específicos deste trabalho: (i) propor características extraídas de imagens que possam ser utilizadas para a classificação de nudez; (ii) avaliar o impacto do uso de tais características na eficácia do sistema proposto; (iii) avaliar o impacto da escolha do espaço de cor na eficácia da classificação de pele e de nudez; e (iv) avaliar a eficácia de redes neurais artificiais para solução dos problemas de classificação encontrados.

### 1.4 Ferramentas de implementação e pseudocódigo apresentado

Para implementação integral dos conceitos e algoritmos propostos neste trabalho foi utilizado o conjunto de ferramentas intitulado MATLAB®. Deste conjunto vale ressaltar que foram utilizadas as ferramentas de construção, treinamento e simulação de redes neurais, manipulação de matrizes, conversão entre espaços de cores e detecção de faces.

No desenvolvimento deste documento são apresentados algoritmos para melhor clareza dos métodos empregados. Estes algoritmos são apresentados em pseudocódigo com sintaxe baseada na própria linguagem do MATLAB e na linguagem C.

### 1.5 Apresentação do manuscrito

No Capítulo 2 é apresentada breve explanação sobre as ferramentas de processamento digital de imagens necessárias à compreensão deste trabalho, como o conceito de imagem digital e o funcionamento de operações morfológicas.

No Capítulo 3 são exploradas técnicas de reconhecimento de padrões utilizadas neste trabalho, como funções discriminantes lineares e redes neurais artificiais. Também é discutido o método



utilizado para medir a eficácia dos classificadores aqui propostos, a *F measure*.

No Capítulo 4 é apresentado o estudo realizado com relação à segmentação por cor de pele utilizando diversos espaços de cores diferentes. A necessidade deste estudo se dá por diversos trabalhos anteriores encontrados indicarem diferentes espaços como alternativas viáveis e com eficácia similar.

No Capítulo 5 é descrito o método proposto para detecção de nudez. No Capítulo 6 são apresentados o conjunto de testes, e os resultados obtidos, bem como descrita a metodologia utilizada para realização dos testes. Finalmente, no Capítulo 7, é realizada avaliação geral do trabalho em relação aos objetivos aqui estabelecidos e à sua contribuição.

## Capítulo 2

# Processamento Digital de Imagens

Processamento Digital de Imagens é uma especialização do Processamento Digital de Sinais cujo sinal processado é uma imagem digital. No entanto, embora a visão seja um dos sentidos mais desenvolvidos do ser humano, ainda é limitada a uma pequena faixa do espectro eletromagnético de luz visível. Contudo, equipamentos eletrônicos não possuem a mesma limitação e são capazes de captar mais faixas do espectro eletromagnético, permitindo-nos obter imagens digitais a partir de sinais que geralmente não associamos à visão como, por exemplo, radar ou imagens de raio-x. Por conta das possibilidades existentes, não é tarefa fácil determinar as fronteiras do processamento de digital de imagens, principalmente quando levadas em consideração outras áreas do conhecimento, como visão computacional, por exemplo. Para os fins deste trabalho, consideramos como processamento digital de imagens operações que possuem imagens como entrada e saída ou imagens como entrada e atributos como saída. Os métodos de processamento digital de imagens utilizados neste trabalho são descritos a seguir.

### 2.1 Imagem digital

Uma imagem pode ser definida como uma função bidimensional,  $f(x, y)$ , na qual  $x$  e  $y$  são coordenadas espaciais (formando um plano) e a amplitude de  $f$  em qualquer coordenada  $(x, y)$  é a intensidade ou nível de cinza da imagem naquele ponto [7]. Quando consideramos que existem valores finitos e discretos para  $x$ ,  $y$  e  $f$ , temos o que definimos uma imagem digital.

A Figura 2.1 ilustra a ampliação de uma imagem digital até que seja possível distinguir claramente as fronteiras entre cada ponto  $(x, y)$ . Cada valor de  $f(x, y)$  é equivalente a um nível diferente de cinza, neste caso, no intervalo 0...255, sendo que o valor 0 corresponde à cor preta e 255, à cor branca. Cada ponto  $(x, y)$  da imagem é também denominado *pixel*. Vale notar que, quando exibida em uma tela na qual muitos *pixels* estão dispostos em área pequena, uma imagem digital – discreta – dá a impressão de continuidade, sem que seja possível perceber as fronteiras entre cada ponto. É importante ressaltar que a representação utilizada foi de 8 *bits* por *pixel*, sendo a imagem obtida por meio do processo de amostragem e quantização em 8 *bits*.

Imagens coloridas, conforme discutido mais à frente, são conjuntos de múltiplas funções bidi-

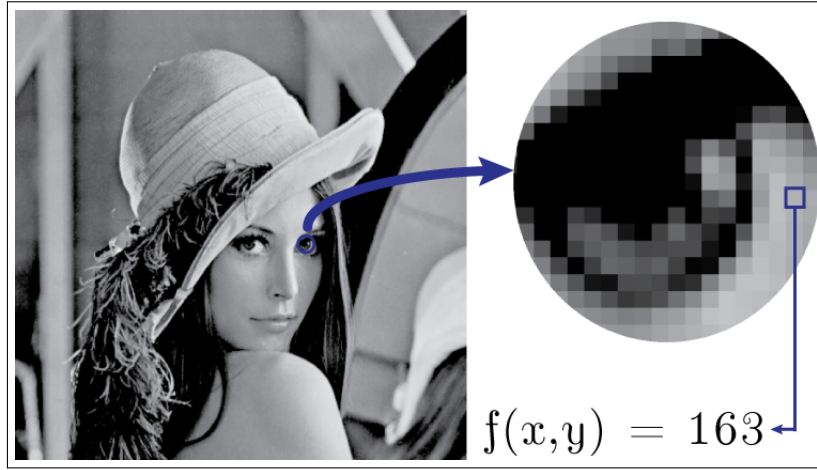


Figura 2.1: Representação digital de uma imagem.

mensionais discretas, cada uma representando um plano do espaço de cor utilizado. Por exemplo, no caso do espaço  $RGB$ , há uma função como esta para representar cada um dos componentes  $R$  (vermelho),  $G$  (verde) e  $B$  (azul).

## 2.2 Operações Morfológicas

Operações morfológicas em imagens dizem respeito às formas encontradas na imagem, como contornos, regiões ou esqueletos. Nesta seção são discutidas as operações morfológicas essenciais a este trabalho.

### 2.2.1 Dilatação

A dilatação é uma operação morfológica primária e serve de base para diversas operações mais complexas utilizadas em processamento digital de imagens. Dado que  $A$  e  $B$  são ambos elementos pertencentes a  $\mathbb{Z}^2$ , a dilatação de  $A$  por  $B$  é denotada por  $A \oplus B$  e definida conforme a equação 2.1.  $B$  é também chamado de elemento estruturante.

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\} \quad (2.1)$$

Esta equação baseia-se na obtenção da reflexão de  $B$  em sua origem e no deslocamento dessa reflexão por  $z$  [7]. Portanto, a dilatação de  $A$  por  $B$  é o conjunto de todos os deslocamentos  $z$  tal que  $\hat{B}$  e  $A$  sobreponham-se em ao menos um ponto. Podemos, portanto, reescrever esta equação como

$$A \oplus B = \{z | [(\hat{B})_z \cap A] \subseteq A\}. \quad (2.2)$$

A dilatação está ilustrada pela Figura 2.2a, na qual o quadrado azul escuro foi dilatado pelo elemento estruturante círculo semitransparente cujo ponto de referência é indicado por um círculo vermelho em seu centro. O resultado da dilatação é o quadrado com bordas arredondadas azul

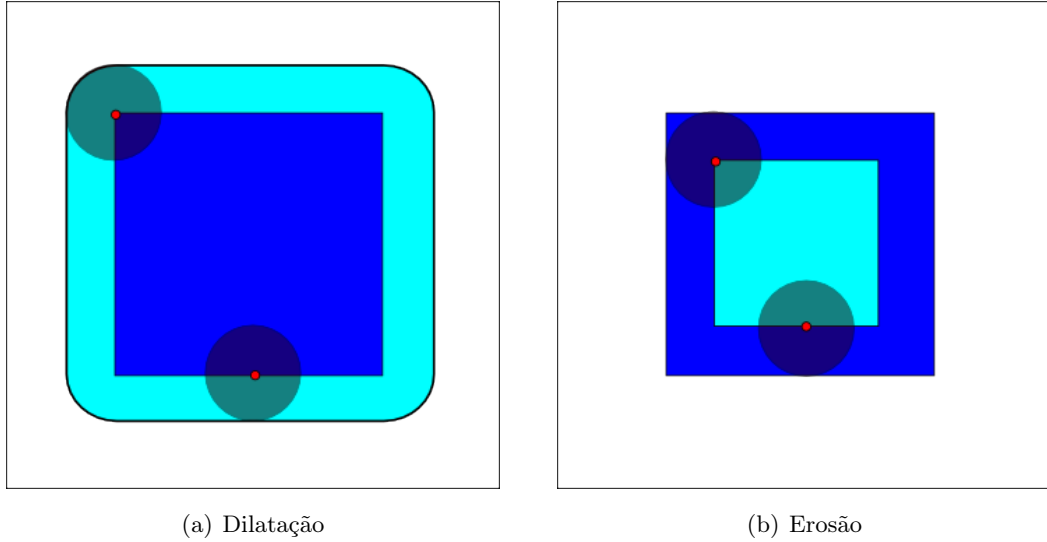


Figura 2.2: Dilatação e erosão de um quadrado azul escuro por um círculo.  
 Figuras disponíveis em [wikimedia.org](https://commons.wikimedia.org/) sob licença GFDL.

claro. Vale notar que o quadrado resultante possui área total preenchida, inclusive a encoberta pelo quadrado azul escuro na Figura 2.2a.

### 2.2.2 Erosão

Para os conjuntos  $A$  e  $B$  em  $\mathbb{Z}^2$  a erosão de  $A$  por  $B$ , denotada por  $A \ominus B$  é definida como

$$A \ominus B = \{z | (B)_z \subseteq A\}. \quad (2.3)$$

Portanto, a erosão é o conjunto de todos os pontos  $z$  tal que  $B$ , trasladado por  $z$ , está contido em  $A$ . A erosão está ilustrada pela Figura 2.2b, na qual o quadrado azul escuro foi erodido pelo elemento estruturante círculo semitransparente. O resultado da erosão está representado pelo quadrado azul claro.

### 2.2.3 Abertura e Fechamento

Conforme ilustrado pela Figura 2.2, enquanto a dilatação acaba por aumentar as dimensões de objetos na imagem, a erosão faz justamente o contrário, encolhendo objetos da imagem. Nesta seção discutimos duas operações morfológicas muito importantes que possuem como base a erosão e a dilatação: a abertura e o fechamento. A *abertura* geralmente suaviza o contorno de um objeto, une pequenas reentrâncias e elimina projeções finas. O *fechamento* também acaba por suavizar seções de contornos porém, ao contrário da abertura, tende a unir pequenas descontinuidades, elimina pequenas falhas e preenche falhas no contorno.

A abertura de um conjunto  $A$  pelo elemento estruturante  $B$ , denotada  $A \circ B$ , é definida como

$$A \circ B = (A \ominus B) \oplus B. \quad (2.4)$$

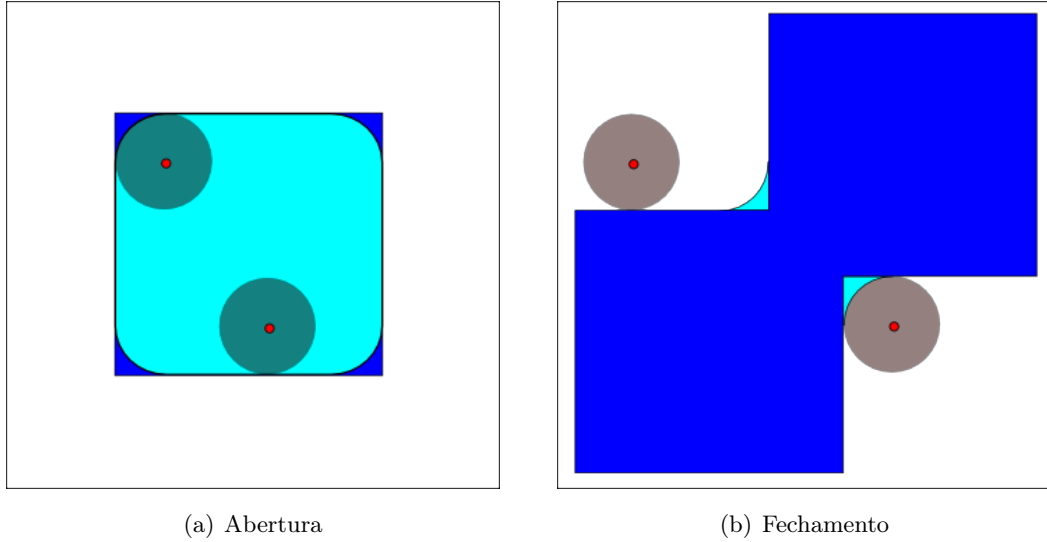


Figura 2.3: Abertura e fechamento de formas em azul escuro por um círculo.  
 Figuras disponíveis em [wikimedia.org](http://wikimedia.org) sob licença GFDL.

Portanto, a abertura de  $A$  por  $B$  nada mais é que a erosão de  $A$  por  $B$  seguida da dilatação do resultado por  $B$ . Da mesma maneira, o fechamento de  $A$  por  $B$ , denotado por  $A \bullet B$ , é definido por

$$A \bullet B = (A \oplus B) \ominus B. \quad (2.5)$$

O fechamento consiste na realização das mesmas operações porém na ordem inversa, ou seja, o fechamento de  $A$  por  $B$  consiste na dilatação de  $A$  por  $B$  seguida da erosão do resultado por  $B$ .

A operação de abertura, ilustrada na Figura 2.3a, possui interpretação geométrica simples, podendo ser traduzida na união de todos os deslocamentos de  $B$  que estiverem completamente contidos em  $A$ . No caso da Figura 2.3a, todos os deslocamentos do círculo de centro vermelho contidos no quadrado azul escuro, resultando no quadrado azul claro.

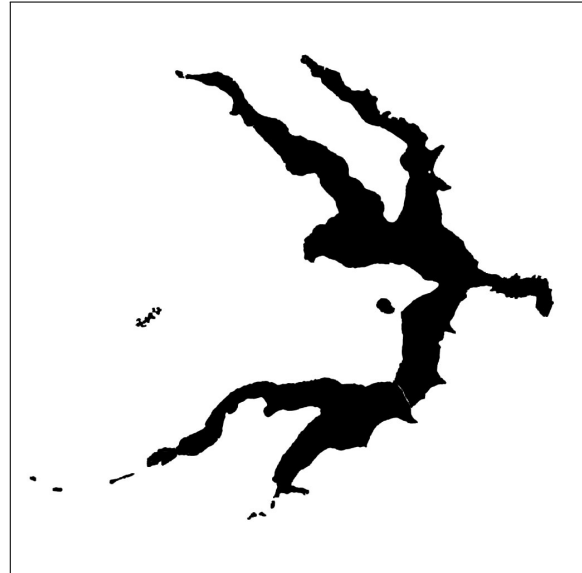
Já a operação de fechamento pode ser interpretada de maneira similar, porém com o deslocamento do elemento estruturante  $B$  pela área externa ao conjunto  $A$ . Geometricamente, um ponto  $x$  é um elemento de  $A \bullet B$  se e apenas se  $(B)_z \cap A \neq \emptyset$  para todo deslocamento  $(B)_z$  que contém  $x$ . A Figura 2.3b representa o fechamento do objeto azul escuro pelo círculo de centro vermelho resultando área total em dois tons de azul.

## 2.3 Segmentação de imagens

A segmentação divide uma imagem em suas partes constituintes ou objetos. [7]. Extrapolando os processamentos que tem como entrada e também saída apenas imagens, a segmentação é um importante passo em direção à extração de informações a partir de imagens. Aplicações que extraem informações de imagens são, em geral, mais complexas que as que tem como saída outras imagens. Isto ocorre devido à necessidade do algoritmo extrair de uma representação exata (*e.g.* matrizes de valores inteiros para as cores vermelho, verde e azul) informações subjetivas que nós,



(a) Imagem de satélite de Brasília



(b) Segmentação de corpos d'água em (a)

Figura 2.4: Segmentação de corpos d'água em uma imagem de satélite de Brasília-DF.

Figura 2.4a disponibilizada pelo programa LANDSAT 8 sob licença *Creative Commons*.

seres humanos, atribuímos às imagens.

A segmentação, por vezes, apresenta um desafio intimamente ligado à aplicação em questão. Por conta de cada domínio ter amostras completamente diferentes (e.g. fotos de alta resolução, imagens geradas por computador, *frames* de vídeos), os desafios são igualmente diferentes. A segmentação, em geral, apresenta-se um problema mais difícil quando há pouco contraste entre a área que deseja-se segmentar e o restante da imagem, devido à utilização da variação entre os valores de representação da imagem para segmentação.

No entanto, a correta segmentação de imagens tem grande impacto no resultado final da aplicação, uma vez que os processamentos subsequentes assumem que as partes em que a imagem foi dividida ou os objetos que foram encontrados estão corretos. Por exemplo, em uma aplicação de mapas cujo objetivo é identificar corpos de água em uma imagem de satélite, erros de segmentação podem acarretar três tipos de falhas indesejadas: (i) identificação de áreas que não possuem água como água (falso positivo); (ii) não identificação de áreas que possuem água (falso negativo); e (iii) identificação incorreta das fronteiras dos objetos de água.

A Figura 2.4 exemplifica a segmentação da aplicação descrita no parágrafo anterior. A Figura 2.4a representa a entrada do processamento enquanto a Figura 2.4b representa a saída do segmentador de corpos d'água, sendo estes identificados pela cor preta.

### 2.3.1 Segmentação por cor de pele

A segmentação de imagens por cor de pele é de extrema utilidade, principalmente quando considerado que pode ser o primeiro passo para segmentação de seres humanos (ou partes de)

em imagens e vídeos. Diversas aplicações fazem o uso de segmentação por cor de pele como, por exemplo, detectores e rastreadores de face [8, 9, 10] e classificadores de gestos [11, 12].

No entanto, mesmo sendo extensivamente utilizada para reconhecimento e segmentação, cor não é um fenômeno físico. Cor é um fenômeno perceptivo relacionado com as características espectrais da radiação eletromagnética nos comprimentos de onda que atingem a retina [13]. Utilizar cor como uma característica para detecção de nudez apresenta vários desafios. A representação em cores do corpo humano capturada por uma câmera é influenciada por diversos fatores, como luz ambiente, movimentação de objetos, etc. Além disso, câmeras diferentes produzem valores significativamente diferentes até para o mesmo corpo sob as mesmas condições de iluminação. Em último lugar, cores de pele diferem entre indivíduos. A fim de utilizar cor de pele como uma característica, precisamos resolver estes problemas.

O primeiro passo para a segmentação por cor de pele é a definição de um modelo de cor de pele. Este modelo é a definição das características que, quando satisfeitas, determinarão que um *pixel* é de pele. Diversos modelos de cor de pele foram propostos e estudados, sendo sua principal diferença o espaço de cor utilizado para proposição do modelo. Entre os espaços de cor já estudados podemos citar *RGB*, *RGB normalizado*, *HSV*, *YCbCr* e *CIELab*. Além do espaço de cores utilizado, também diferem os métodos utilizados para definição, a partir das informações de cor no espaço escolhido, se determinada cor é pele ou não. Dentre os métodos já desenvolvidos podemos citar limiarização de valores, função gaussiana bidimensional [14], correlação entre espaços de cor, modelo de probabilidades bayesianas, *lookup tables* e redes neurais artificiais.

## 2.4 Espaços de Cor

Conforme discutido na Seção 2.1, imagens digitais são matematicamente funções bidimensionais discretas. No entanto, a experiência sensorial de visão do ser humano acontece em cores, e níveis de cinza não são suficientes para representá-la. Portanto, com base na representação de imagem digital já conhecida, foram propostos diversos modelos de representação de cores, cada um com suas premissas e objetivos específicos.

Apesar de bastante diferentes, todas as representações do espaço de cores exploradas nesta seção baseiam-se na representação de imagem digital discutida anteriormente. Enquanto a imagem em níveis de cinza é representada por uma função bidimensional  $f(x, y)$ , uma imagem colorida será representada por duas ou três funções  $f(x, y)$ , uma para cada componente de cor.

### 2.4.1 *RGB*

O espaço de cores *RGB* é amplamente utilizado como representação padrão de imagens coloridas. Sua popularidade está fortemente ligado ao tipo de dispositivos que utilizamos para visualização destas imagens. Diferentes tecnologias para projeção de imagens (i.e. CRT, LCD, LED, projetores CRT, LCD, DLP e LCOS) todos utilizam combinações de diferentes níveis das três cores primárias vermelho (R), verde (G) e azul (B) para produzir todas as cores no espectro visível, pelo

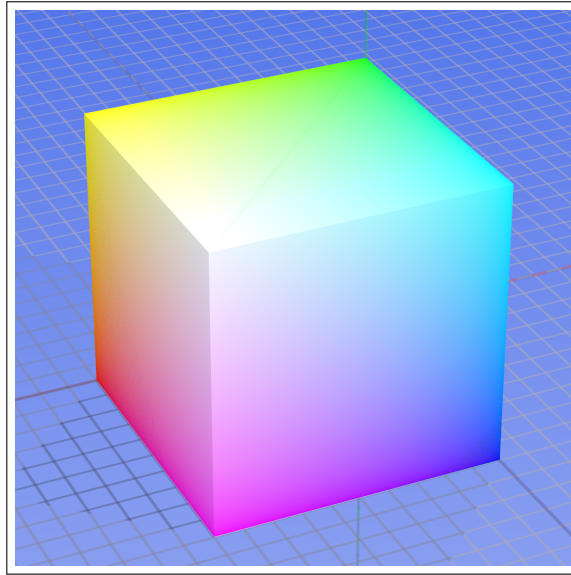


Figura 2.5: Representação do espaço de cores *RGB*.  
Imagem disponível em [commons.wikimedia.org](https://commons.wikimedia.org) sob licença CC BY-SA 3.0.

fado das cores vermelho, verde e azul serem cores primárias complementares aditivas.

Portanto, a representação em *RGB* especifica cores com valores de intensidade para cada cor primária (vermelho, verde e azul), geralmente com valores no intervalo de 0 a 255, para representação de cores em 24 *bits*. Neste caso, por exemplo a combinação das cores puras vermelho ( $\{R, G, B\} = \{255, 0, 0\}$ ) e verde ( $\{0, 255, 0\}$ ) resulta na cor amarelo ( $\{255, 255, 0\}$ ). Nesta representação, se todos os componentes possuírem intensidade máxima ( $\{255, 255, 255\}$ ), a cor branca é obtida. Por outro lado, com a intensidade mínima em todos os componentes obtém-se a cor preta ( $\{0, 0, 0\}$ ).

A Figura 2.5 é uma representação do espaço de cores *RGB* na qual os eixos  $x$ ,  $y$  e  $z$  representam a intensidade das cores vermelho, azul e verde, respectivamente.

#### 2.4.2 *nRGB*

A representação *nRGB* nada mais é que o espaço *RGB* normalizado. O processo de normalização tem como principal objetivo o aumento de robustez a variações causadas por luzes e sombras. Apesar de simples, a normalização é uma maneira eficaz de obter o resultado desejado.

O espaço *nRGB* ( $\{r, g, b\}$ ) tem suas componentes calculadas a partir das intensidades  $R$ ,  $G$  e  $B$  componentes do espaço de cores *RGB*. Cada componente *nRGB* representa a intensidade da componente correspondente no espaço *RGB* dividido pela soma das intensidades das componentes do espaço *RGB*, conforme especificado nas Equações 2.6, 2.7 e 2.8.

$$r = \frac{R}{R + G + B} \quad (2.6)$$



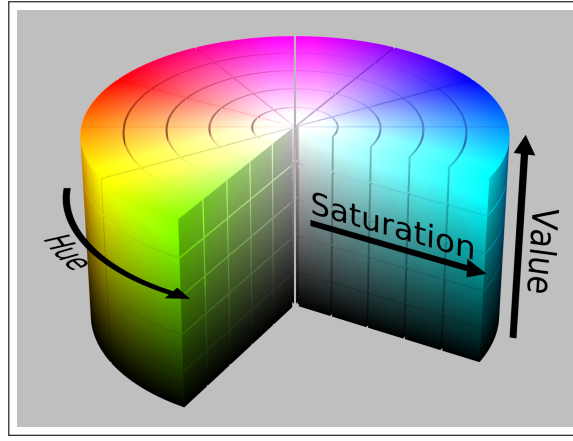


Figura 2.6: Representação do espaço de cores HSV.  
Imagem disponível em commons.wikimedia.org sob licença CC BY-SA 3.0.

$$g = \frac{G}{R + G + B} \quad (2.7)$$

$$b = \frac{B}{R + G + B} \quad (2.8)$$

Como consequência, vale notar que a soma de todas as componentes do espaço  $nRGB$  será sempre igual a 1, conforme a Equação 2.9

$$r + g + b = 1 \quad (2.9)$$

### 2.4.3 $nRG$

Conforme descrito pela Equação 2.9, a soma das componentes do espaço  $nRGB$  é conhecida. Portanto, podemos omitir a componente  $b$  sem prejuízo de informações, assumindo que pode ser derivada das componentes  $r$  e  $g$ , conforme as Equação 2.10.

$$b = 1 - r - g \quad (2.10)$$

Desconsiderando a componente  $b$  do espaço  $nRGB$  obtemos o espaço  $nRG$ , com dimensionalidade reduzida para duas dimensões sem prejuízo de informação, dessa maneira simplificando os métodos de segmentação.

### 2.4.4 $HSV$

O espaço de cores  $HSV$  (*Hue*, *Saturation* e *Value*) surgiu da necessidade do usuário especificar propriedades de cor numericamente. Este espaço descreve cores de maneira intuitiva, baseado na definição pelos artistas de matiz, saturação e tom. A propriedade *Hue* (matiz) define a cor

dominante (como vermelho, verde, roxo e amarelo), enquanto a *Saturation* (saturação) mede o quanto colorida a cor é em relação ao seu brilho. Já a propriedade *Value* (valor) está relacionada à luminância da cor.

Por ser baseada na percepção do ser humano sobre as cores, o espaço *HSV* é popular entre trabalhos de segmentação de cor de pele. Duas propriedades interessantes deste espaço de cor foram destacadas em [15]: (i) invariância a reflexos de fontes de luz branca; e (ii) invariância a mudanças na orientação de superfícies foscas em relação a fontes de luz.

No entanto, algumas características do *HSV* são indesejáveis, como a descontinuidade da característica *Hue* e o brilho ser computado a partir de diferentes propriedades [16], o que vai de encontro às propriedades da percepção de cor.

O espaço de cor *HSV* é computado a partir do espaço *RGB* por meio das Equações 2.11, 2.12 e 2.13. A Figura 2.6 provê uma representação visual do espaço de cores *HSV*.

$$H = \arccos \frac{\frac{1}{2}((R - G) + (R - B))}{\sqrt{((R - G)^2 + (R - B)(G - B))}} \quad (2.11)$$

$$S = 1 - 3 \frac{\min(R, G, B)}{R + G + B} \quad (2.12)$$

$$V = \frac{1}{3}(R + G + B) \quad (2.13)$$

#### 2.4.5 $YC_bC_r$

O espaço de cores  $YC_bC_r$  é uma codificação não linear do sinal *RGB* bastante utilizada por sistemas de televisão europeus. Neste espaço, a luminância é representada pelo componente *Y*, este calculado a partir de combinação dos componentes *R*, *G* e *B* [16]. A cor é representada por dois componentes de diferença de cor,  $C_b$  e  $C_r$ , formados a partir da subtração de *Y* dos componentes vermelho e azul do espaço *RGB*, conforme descrito nas Equações 2.14, 2.15 e 2.16. A Figura 2.7 ilustra a distribuição de cores no espaço  $C_bC_r$  para diferentes valores de *Y*.

$$Y = 0.299R + 0.587G + 0.114B \quad (2.14)$$

$$C_b = B - Y \quad (2.15)$$

$$C_r = R - Y \quad (2.16)$$

A simplicidade da transformação e o isolamento da luminância no componente *Y* tornam este espaço de cor muito atraente para modelagem de cor de pele. Segundo Fleck et al. [17, 18], a cor da pele humana é uma combinação de sangue (vermelho) e melanina (amarelo e marrom) e, portanto,

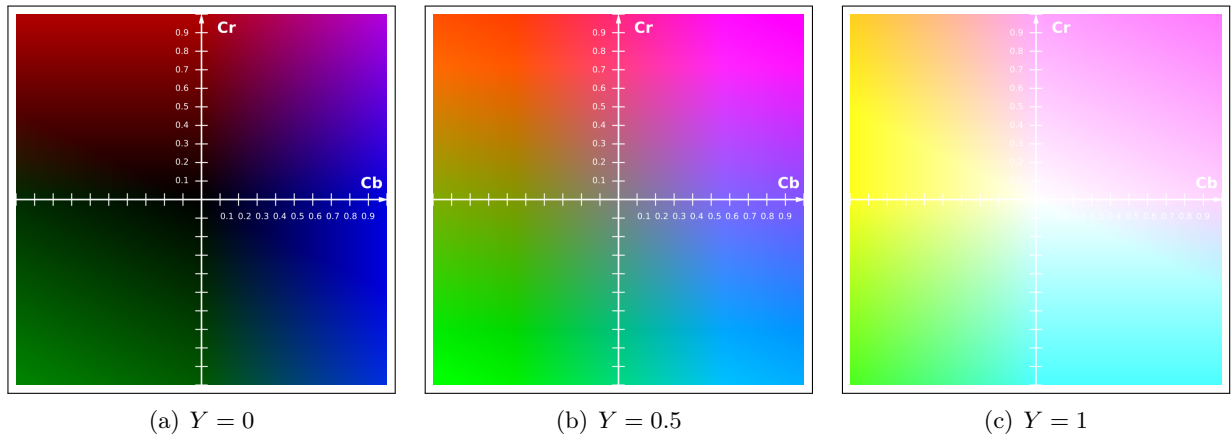


Figura 2.7: Distribuição de cores no espaço  $C_b C_r$  para diferentes valores de  $Y$   
 Imagem disponível em [commons.wikimedia.org](https://commons.wikimedia.org) sob domínio público.

possui uma quantidade limitada de matizes. Considerando esta descrição da cor de pele, pode-se descartar a luminância e utilizar apenas a cromaticidade (esta descrita em termos de  $C_b$  e  $C_r$ ).

## Capítulo 3

# Reconhecimento de Padrões

Frequentemente subestimamos nossa capacidade como seres humanos de reconhecer faces, entender palavras quando pronunciadas, ler caracteres escritos, identificar um objeto apenas pelo tato ou decidir se o leite está estragado por seu cheiro. No entanto, essas tarefas de reconhecimento de padrões constituem processos bastante complexos. Reconhecimento de padrões é o ato de assimilar dados crus e tomar uma ação baseada na “categoria” do padrão [19] e, apesar de subestimada, tem sido essencial para nossa sobrevivência como espécie. Ao longo de nossa jornada evolutiva desenvolvemos sistemas neurais e cognitivos altamente complexos para lidar com essas tarefas e o esforço empregado no desenvolvimento de algoritmos que habilitem sistemas computacionais a executá-las demonstra quão complexas podem ser.

Podemos citar inúmeras aplicações para máquinas que desempenham reconhecimento de padrões, como reconhecimento automático de fala, sistemas biométricos (e.g. reconhecimento de impressões digitais, padrões de voz, face, íris, postura, digitação), reconhecimento óptico de caracteres (OCR), inspeção visual de linhas de produção, entre outras. Portanto, é natural o ímpeto de desenvolver máquinas que reconheçam padrões. Ainda que durante a resolução desses problemas podemos efetivamente aprofundar nosso entendimento sobre mecanismos naturais de reconhecimento de padrões, por vezes utilizamos o conhecimento já adquirido para desenvolver nossas soluções, como modelar o mecanismo da fala a partir da anatomia já conhecida, por exemplo.

Neste trabalho o objetivo almejado é justamente projetar e desenvolver um algoritmo que seja capaz de reconhecer automaticamente a presença de nudez em imagens digitais. A tarefa em questão pode ser descrita como separar imagens (representação digital de estímulos captados por sensor óptico) em dois grupos: *(i) imagens que contém nudez*; e *(ii) imagens que não contém nudez*. Embora essa tarefa possa ser simples para um ser humano, passa a não ser tão trivial quando consideramos que depende de nossa capacidade de reconhecimento de cores de pele e de partes do corpo humano aliada ao nosso julgamento subjetivo baseado em nossa definição de nudez, o que pode variar entre indivíduos.

Neste capítulo são explorados os métodos de reconhecimento de padrões pertinentes a este trabalho, bem como suas vantagens, desvantagens e desafios, a fim de subsidiar posterior explanação do método proposto.

## 3.1 Aspectos gerais de sistemas de reconhecimento de padrões

### 3.1.1 Modelo

Em tarefas de reconhecimento de padrões, o primeiro passo é a definição dos *modelos* que se pretende classificar. Por exemplo, em uma aplicação cujo objetivo é identificar produtos com deformidades em uma linha de produção, podemos definir dois modelos: *(i)* produtos com deformidades; e *(ii)* produtos sem deformidades. Em seguida, devemos definir as *características* que utilizamos para identificar cada modelo e diferenciá-los um do outro.

No exemplo em questão, tais características podem ser as mais diversas, como a cor do produto, sua forma, a textura de sua superfície, massa, entre outros. Características devem ser escolhidas levando-se em consideração dois princípios: relevância e viabilidade. O primeiro diz respeito à importância da característica para diferenciação dos modelos estabelecidos, ou seja, deve ser observada variação da característica entre os diferentes modelos que se deseja classificar. O segundo diz respeito ao custo de utilização da característica, ou seja, esta deve exigir extração e processamento tecnologicamente possível, de maneira a viabilizar o projeto e desenvolvimento do sistema. Por exemplo, a utilização de espectrometria a laser pode não ser viável em nosso exemplo de linha de produção devido ao seu alto custo e baixa vazão.

### 3.1.2 Pré-processamento e segmentação

Digamos que para este exemplo escolhamos as características cor predominante e forma do produto inspecionado. Operações de *pré-processamento* podem ser utilizadas para simplificar operações subsequentes sem a perda de informação relevante. Mais especificamente, podemos utilizar uma operação de *segmentação* para separar as imagens de cada produto de outros produtos e da esteira da linha de produção. Outras operações de pré-processamento também podem ser realizadas, como equalização de histograma ou ajustes de luminosidade, a fim de obter resultados mais uniformes, por exemplo.

É importante notar que todas as operações realizadas nesta etapa tem como objetivo a simplificação das operações subsequentes, mais especificamente, as operações de *extração de características*.

### 3.1.3 Extração de características e amostras de treinamento

Uma vez definidos os modelos e suas respectivas características, após realizadas as operações de pré-processamento, partimos à extração das características escolhidas. Em nosso exemplo, já com a imagem de cada produto corretamente segmentada, podemos fazer uma simples média de valores para determinar a cor predominante da amostra. Podemos também definir o contorno da amostra, o descrevendo em termos de dimensões, arestas e ângulos, por exemplo.

Uma vez realizada a extração de características, é essencial a definição de um conjunto de *amostras de treinamento*, amostras representativas dos indivíduos dos diferentes modelos que buscamos

identificar e reconhecer. Este conjunto será utilizado para validação das hipóteses previamente estabelecidas e realização de ajustes no sistema proposto.

### 3.1.4 Custo do erro e teoria de decisão

Até agora assumimos que o custo do erro é igual para ambos os casos: classificar um produto adequado como inadequado é tão prejudicial quanto o contrário. Embora em muitas situações o custo seja de fato equivalente, isso não é sempre verdade. Por exemplo, podemos considerar que a chegada de um produto defeituoso nas mãos de um cliente é muito mais prejudicial que o descarte e reciclagem de um produto adequado. Nesse caso devemos ajustar nossa fronteira de decisão de maneira a garantir que produtos defeituosos não sejam identificados como adequados.

Com o custo do erro bem definido, podemos ajustar nosso sistema e escolher a fronteira de decisão ótima de maneira a minimizar tal custo. Esta é a principal tarefa da *teoria de decisão*, talvez a mais importante parte do reconhecimento de padrões.

A teoria de decisão é justamente o método utilizado para, a partir das características extraídas e do que é conhecido a partir das amostras de treinamento, decidir qual dos modelos descreve melhor cada amostra analisada pelo sistema.

Nas Seções 3.2 e 3.3 são descritas teorias de decisão úteis a este trabalho.

## 3.2 Funções discriminantes lineares

No caso da classificação de padrões, nosso objetivo é separar, com o menor erro possível dois ou mais grupos de indivíduos. Ou seja, dado que conhecemos algumas amostras de cada grupo, podemos determinar a qual grupo pertence um novo indivíduo que nos seja apresentado.

Na Figura 3.1 temos uma ilustração da distribuição de indivíduos de dois grupos distintos (pontos pretos e pontos brancos) no espaço bidimensional de duas características exploradas. Uma característica distribuída na no eixo das abscissas e a outra, no eixo das ordenadas. Cada indivíduo está representado como um ponto na combinação de suas duas características. Considerados os parâmetros aceitáveis de erro, podemos definir que a reta pontilhada é o que separa os dois grupos identificados. Indivíduos acima da reta serão classificados como pontos brancos enquanto indivíduos abaixo da reta serão classificados como pontos pretos.

Quando é assumido que as funções que separam os grupos de indivíduos são conhecidas, a teoria de decisão de *funções discriminantes* pode apresentar bons resultados de classificação. Para efeitos deste trabalho consideraremos as funções discriminantes lineares em  $x$  ou lineares em um dado conjunto de funções de  $x$ .

Funções discriminantes possuem diversas propriedades analíticas convenientes e simplificam muitas operações que são bastante complexas em outros cenários. Especialmente quando é possível sacrificar um pouco do desempenho do sistema proposto, funções lineares podem ser a base de classificadores simples porém eficientes. Na maioria dos casos, classificadores como estes são um

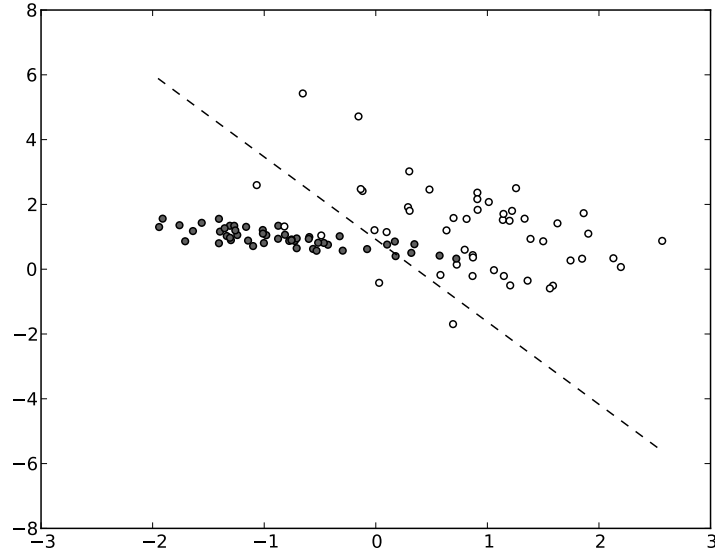


Figura 3.1: Distribuição de indivíduos de dois grupos distintos no espaço de suas características. Figura disponível em [wikimedia.org](http://wikimedia.org) sob domínio público.

excelente ponto de partida, estabelecendo um limite inferior que deva ser superado.

### 3.2.1 Funções discriminantes lineares para classificadores com duas categorias

Nesta subseção discutiremos o caso de um classificador que utiliza funções discriminantes lineares para classificar indivíduos pertencentes a dois grupos distintos. A Figura 3.1 ilustra um exemplo simplificado deste caso.

Podemos descrever a função discriminante linear como uma combinação de componentes de  $x$

$$g(x) = w^t x + w_0 \quad (3.1)$$

na qual  $w$  é o *vetor de pesos* e  $w_0$  é o *viés* ou *peso limiar*. Isto posto, um classificador linear com duas categorias utiliza a seguinte teoria de decisão: grupo  $w_1$  se  $g(x) > 0$  e  $w_2$  se  $g(x) < 0$ . Portanto,  $x$  é classificado como  $w_1$  caso o produto interno  $w^t x$  supere o limiar  $-w_0$  ou como  $w_2$  caso contrário. No caso  $g(x) = 0$ ,  $x$  pode ser atribuído a qualquer categoria.

A equação  $g(x) = 0$  define a superfície de decisão que separa pontos pertencentes a  $w_1$  de pontos pertencentes a  $w_2$ . Quando  $g(x)$  é linear, a superfície de decisão é um hiperplano, ou seja, possui  $n - 1$  dimensões no espaço amostral de  $n$  características (ou dimensões). Por definição, um hiperplano divide o espaço em que está contido em dois subespaços, neste caso, cada espaço dos elementos pertencentes a um dos grupos. No exemplo da Figura 3.1 a superfície de decisão é a reta pontilhada, um hiperplano do espaço bidimensional de características.

Funções discriminantes lineares apresentam uma solução simples para o problema de decisão no reconhecimento de padrões. Apesar de lograr bons resultados em casos específicos, por vezes o

problema encontrado não é solucionável por meio de uma combinação de funções lineares. Ainda mais, há o problema de determinação das funções lineares utilizadas, mais especificamente seus parâmetros, que possui diversas soluções, como descendente gradiente, descendente newtoniano, perceptron incremental, descendente fixo incremental, algoritmo balanceado de Winnow, entre outros [19]. Tais algoritmos apresentam melhores resultados para casos específicos e, portanto, a tarefa de determinar qual é o mais adequado em dada situação não é simples.

### 3.3 Redes Neurais Artificiais

Apesar do método de Funções Discriminantes Lineares apresentar uma boa solução para problemas reais, especialmente por meio da utilização de algoritmos para redução de erros, por vezes a classe de soluções que podem ser obtidas por tal método (hiperplanos discriminantes) não são gerais o suficiente para aplicações em que o menor erro não é atingível com o uso de discriminantes lineares.

No entanto, com a escolha certa de funções não lineares  $\varphi$  é possível obter decisões arbitrárias, em especial as que nos levam ao erro mínimo. Naturalmente, o desafio neste caso é determinar as funções mais adequadas. É necessário, portanto, uma maneira de aprender a não-linearidade a partir das amostras de treinamento. Esta é a abordagem utilizada pelas Redes Neurais Artificiais (RNAs). Na prática, RNAs implementam um conjunto de discriminantes não lineares e, portanto, não há nada de mágico em seu funcionamento. Para efeitos deste trabalho estamos interessados nas RNAs de múltiplas camadas do tipo *feedforward*, que terão seu funcionamento descrito e serão utilizadas como classificadores.

O correto uso de RNAs simplifica problemas de reconhecimento de padrões por prover método para determinar automaticamente os parâmetros de configuração. Ao menos em teoria, RNAs fornecem solução ótima para qualquer problema de classificação. Nesta Seção discutiremos suas características e formas de uso.

#### 3.3.1 Neurônio

As RNAs baseiam-se no princípio de dividir um problema complexo e processá-lo paralelamente em diversas unidades extremamente simples de processamento. Tal unidade é denominada neurônio, por ser um modelo computacional do funcionamento de um *neurônio biológico*.

O neurônio biológico, ilustrado pela Figura 3.2, é uma célula que possui  $n$  ligações de entrada e apenas uma de saída. A partir dos estímulos recebidos em suas ligações de entrada, é emitido ou não estímulo para sua ligação de saída. Portanto, podemos considerar que o neurônio é um elemento simples de processamento que, a partir de  $n$  entradas, emite ou não uma saída. A *função de ativação*, que determina a saída de um neurônio a partir de suas entradas, é o processamento realizado por ele. Esta é também conhecida como função de transferência.

Tal interpretação de um neurônio biológico nos permite facilmente traduzir seu funcionamento para um modelo de *neurônio computacional*. O primeiro modelo matemático de um neurônio foi



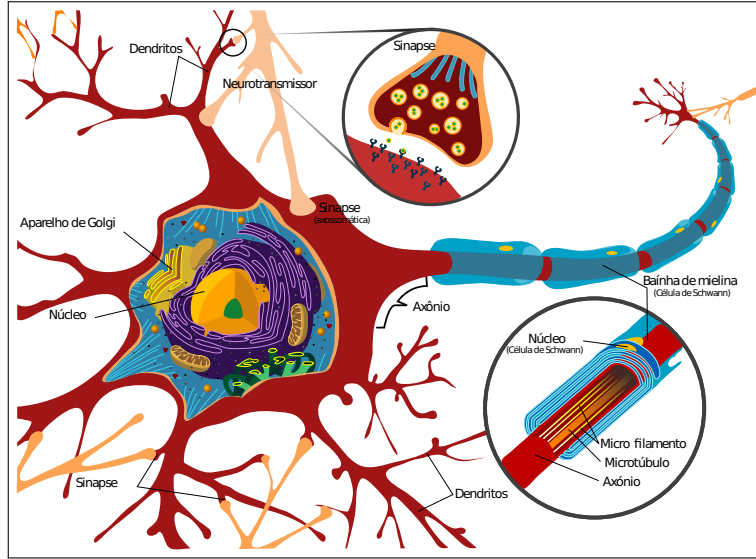


Figura 3.2: Representação de neurônio biológico.  
Adaptado de wikimedia.org sob domínio público.

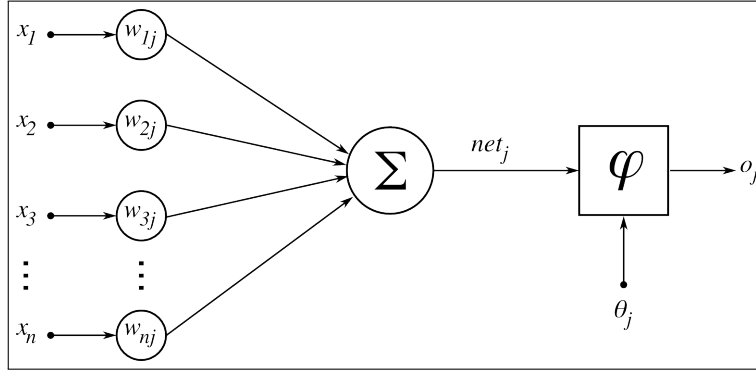


Figura 3.3: Representação de neurônio MCP.  
Disponível em wikimedia.org sob licença GDFL.

proposto por McCulloch e Pitts em 1943 [20] e é conhecido como MCP. Sua proposta é simples, porém bastante eficaz, e até hoje serve de base para as redes neurais que utilizamos.

O modelo MCP, ilustrado na Figura 3.3, define o resultado da computação do neurônio  $j$  como

$$o_j = \varphi\left(\sum_{k=1}^n w_{kj}x_k, \theta_j\right), \quad (3.2)$$

sendo que a função de ativação  $\varphi$  calcula o resultado a partir da soma de todas as entradas  $x_n$  multiplicadas por seus respectivos pesos  $w_{nj}$  e o *limiar* estabelecido  $\theta_j$ .

A partir deste modelo podemos explorar algumas funções de ativação  $\varphi$  que resultam em comportamentos ligeiramente diferentes para nosso modelo de neurônio computacional. A fim de simplificar a notação, deste ponto em diante utilizaremos o símbolo  $\theta$  para denotar a soma das entradas do neurônio multiplicadas por seus respectivos pesos, de maneira que

$$\theta = \sum_{k=1}^n w_{kj}x_k. \quad (3.3)$$

### 3.3.1.1 Funções de ativação degrau unitária e signum

A função degrau unitária é a mais simples e foi a proposta originalmente no modelo de neurônio MCP. Esta é definida como

$$\varphi(\theta, \theta_j) = \begin{cases} 1 & , \theta > \theta_j \\ 0 & , \theta \leq \theta_j \end{cases} . \quad (3.4)$$

Utilizando esta função de transferência, o neurônio é ativado quando a soma das entradas  $x_n$  multiplicadas por seus respectivos pesos  $w_{nj}$  supera o limiar  $\theta_j$ . Caso contrário, o neurônio permanece desativado. A Figura 3.4a ilustra graficamente o comportamento da função degrau.

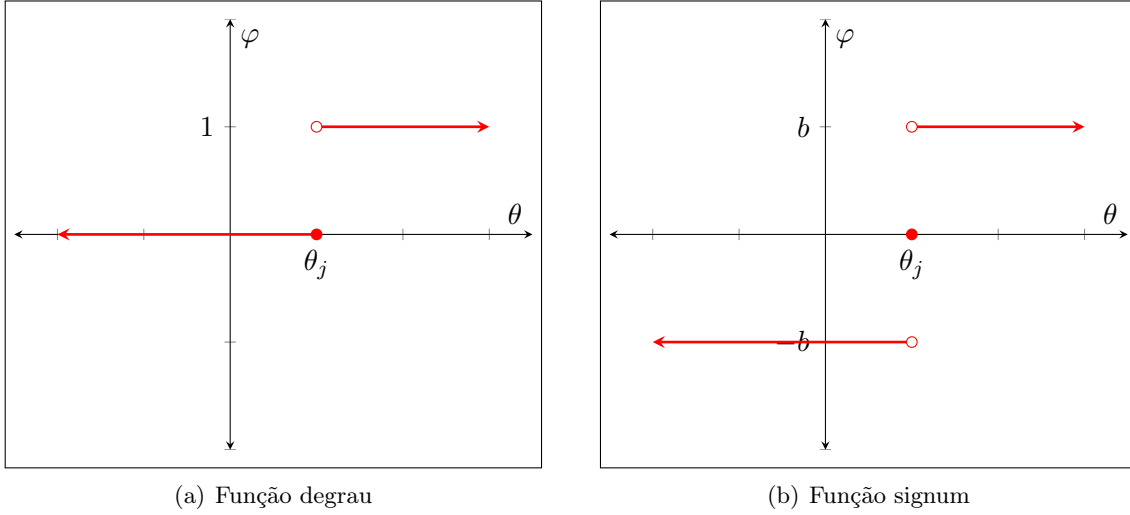


Figura 3.4: Funções de ativação degrau e signum.

A função signum  $sgn(x)$  é bastante semelhante à degrau. No entanto, ela possui amplitude definida pela constante  $b$ . A função de ativação signum pode ser definida como

$$\varphi(\theta, \theta_j) = b \operatorname{sgn}(\theta - \theta_j) = \begin{cases} b, & \theta > \theta_j \\ 0, & \theta = \theta_j \\ -b, & \theta < \theta_j \end{cases} . \quad (3.5)$$

### 3.3.1.2 Funções de ativação log-sigmóide e tangente hiperbólica

A função sigmóide  $S(x)$  também é utilizada como função de ativação. Neste caso, a saída do neurônio é determinada pela função

$$\varphi(\theta, \theta_j) = S(\theta - \theta_j) = \frac{1}{1 + e^{\alpha(\theta - \theta_j)}} . \quad (3.6)$$

A função sigmóide resulta em uma transição mais suave entre o estado zero e o estado um do neurônio quando comparada à função degrau. Além disso, esta função também apresenta valores intermediários, entre zero e um, deixando de ser um neurônio com ativação binária. Durante o treinamento da RNA, o parâmetro  $\alpha$  da função de transição também pode ser ajustado, resultando

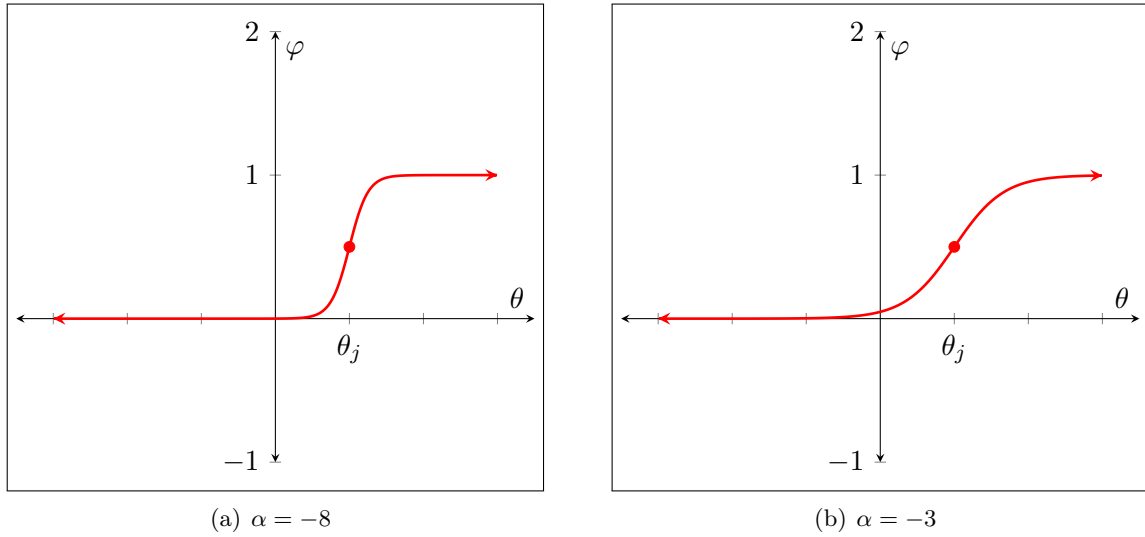


Figura 3.5: Função de ativação sigmóide com diferentes valores de  $\alpha$ .

em uma transição mais suave ou abrupta da saída do neurônio. A Figura 3.5 ilustra a resposta do neurônio com a função de ativação sigmóide para diferentes valores de  $\alpha$ .

A função tangente hiperbólica  $\tanh z$  assemelha-se à função sigmóide. No entanto, sua amplitude varia entre  $-1$  e  $1$ . A função de ativação da tangente hiperbólica está ilustrada na Figura 3.6 e é definida como

$$\tanh(\theta - \theta_j). \quad (3.7)$$

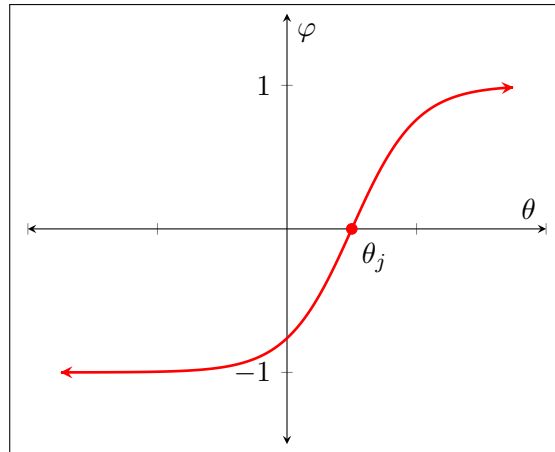


Figura 3.6: Função de ativação tangente hiperbólica.

### 3.3.2 Redes neurais artificiais de múltiplas camadas

Conforme apresentado na Subseção 3.3.1, um neurônio é um elemento de processamento capaz de realizar computações simples. No entanto, o processamento que um neurônio é capaz de realizar geralmente não é suficiente. Neurônios biológicos se combinam para formar estruturas mais complexas, como o cérebro humano, que possui aproximadamente  $10^{11}$  deles.

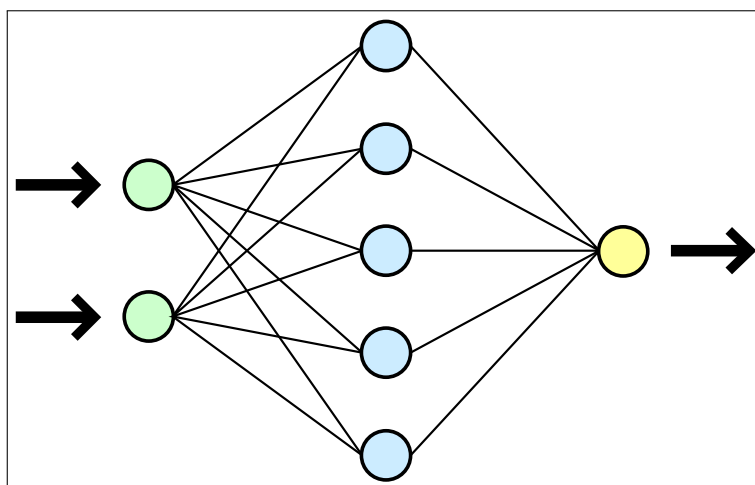


Figura 3.7: Rede neural completamente ligada com dois neurônios de entrada, cinco neurônios na camada escondida e um neurônio de saída.

Disponível em [commons.wikimedia.org](https://commons.wikimedia.org) sob licença Creative Commons.

No cérebro, neurônios se combinam por meio de ligações chamadas sinapses. Estas ligações são criadas e destruídas no processo de desenvolvimento do indivíduo e no processo de aprendizagem. O cérebro é capaz de realizar tarefas complexas como cognição, armazenagem de informação, percepção sensorial e tarefas específicas de reconhecimento de padrões, como o próprio ato de entender o que uma pessoa fala, ler um livro ou identificar uma cadeira por meio da visão.

É natural, a partir do modelo matemático de neurônio definido acima, desenvolver modelo também semelhante ao biológico para as redes que os interligam. Esta é a tarefa das Redes Neurais Artificiais. As RNAs imitam tal comportamento basicamente organizando seus neurônios em uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. A camada de entrada representa as unidades sensoriais do sistema, responsáveis por receber os parâmetros de entrada e repassá-los à frente, sem realizar qualquer processamento. Os neurônios recebem o que chamamos *vetores de entradas*, em qual cada elemento é o valor correspondente de uma das características utilizadas para classificação. O fato da informação trafegar da entrada à saída, sendo sempre processada por camadas cada vez mais próximas à ela, denomina a RNA do tipo *feedforward*, ou seja, alimentada adiante.

O diagrama da Figura 3.7 ilustra uma rede neural completamente ligada que possui dois neurônios na camada de entrada, na cor verde, cinco neurônios na camada escondida, na cor azul, e um neurônio na camada de saída, na cor amarelo. Esta rede possui dois parâmetros de entrada, cada um recebido por seu respectivo neurônio na camada de entrada que o propaga a todos os neurônios da camada escondida. Já o neurônio da camada de saída recebe como entrada a saída de todos os neurônios da camada escondida e realiza a computação final, cujo resultado é também o resultado da própria rede. Neste diagrama, a característica *feedforward* pode ser descrita como o estímulo sempre percorrer a rede da esquerda para a direita sem que haja retroalimentação entre as camadas.

RNAs de múltiplas camadas apresentam três características básicas: (i) a função de ativação

dos neurônios da rede são não-lineares e suaves, ou seja, diferenciáveis em qualquer ponto, como as funções tangente hiperbólica e sigmóide discutidas na Seção 3.3.1.2; (ii) a rede possui uma ou mais camadas de neurônios ocultos, que não fazem parte da entrada ou saída da rede; e (iii) a rede apresenta alto grau de conectividade entre seus neurônios.

### 3.3.2.1 Aprendizado de RNAs com múltiplas camadas

A utilização de RNAs como classificadores baseia-se em sua capacidade de aprender padrões para poder, em seguida, classificá-los. O processo de aprendizado das RNAs consiste em apresentar à rede vetores de características dos indivíduos que se deseja classificar enquanto são ajustados seus parâmetros para minimizar o erro de classificação. Este processo de aprendizado pode ser tanto supervisionado quanto não-supervisionado. No primeiro caso, serão apresentadas também as saídas desejadas para cada conjunto de entrada enquanto no segundo, a decisão dos grupos a serem classificados será tomada de forma automática buscando-se regularidade estatística nos dados de entrada.

Para os fins deste trabalho estamos interessados no modelo de aprendizado supervisionado conhecido como retropropagação de erro (*error backpropagation*). Apesar do termo *backpropagation* ser referência ao próprio algoritmo de cálculo do gradiente do erro, aqui o utilizamos como referência da forma em que o erro é propagado, da saída até os neurônios de entrada a fim de se ajustar os pesos de cada neurônio.

O algoritmo de treinamento utilizado é denominado Levenberg-Marquardt, escolhido principalmente por ser consideravelmente mais rápido quando comparado aos demais. Este algoritmo é o sugerido pela plataforma utilizada justamente por sua alta velocidade de treinamento em redes de tamanho moderado e eficiente implementação no MATLAB. Detalhes sobre seu funcionamento e implementação estão descritos em [21].

## 3.4 Detecção de faces

A detecção de faces é um problema clássico de reconhecimento de padrões que possui diversas aplicações, como identificação biométrica, por exemplo. O problema consiste em, a partir de uma imagem digital, localizar faces humanas presentes. A saída de um classificador de faces é, portanto, um vetor de retângulos que delimitam cada face encontrada.

O detector de faces utilizado neste trabalho é o clássico proposto por Paul Viola e Michael Jones em 2001 [22]. Este detector é adotado como padrão em diversos sistemas e inclusive já é implementado em populares ferramentas de processamento de imagens, como o MATLAB e o OpenCV.

O algoritmo Viola-Jones é um detector de objetos baseado em características Haar e aprendido em que uma função cascata é treinada com um vasto conjunto de testes com faces e sem faces. Detalhes sobre este método fogem do escopo deste trabalho e podem ser encontrados em [22]. Para implementação, é utilizada a padrão fornecida pelo MATLAB.

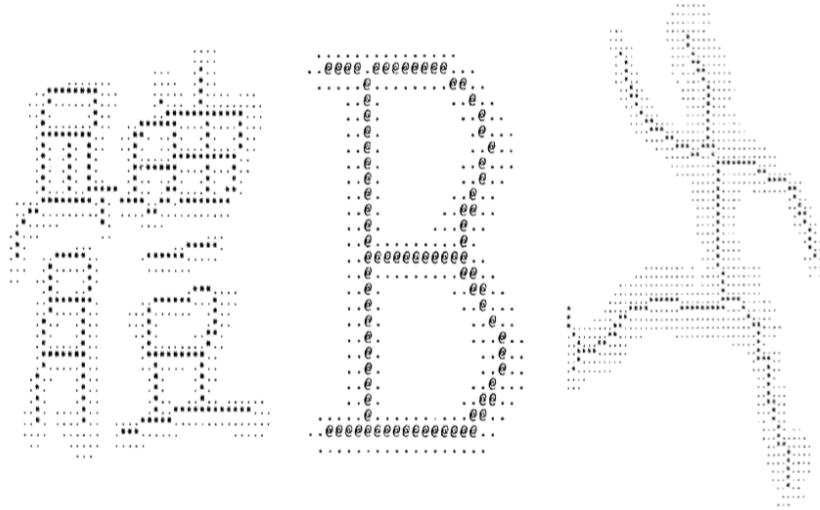


Figura 3.8: Processo de afinamento de imagens binárias.

Figura adaptada de [24].

### 3.5 Reconhecimento de padrões morfológicos

Enquanto nas Seções 3.2, 3.3 e 3.4 anteriores discutimos classificadores, nesta apresentamos operações morfológicas e outras técnicas para extração de características a partir de imagens binárias.

A primeira operação é chamada de *afinamento* ou *esqueletização*. Muito utilizada em operações de reconhecimento de padrões, surgiu pela necessidade de redução da quantidade de informações para processamento nos primórdios da computação porém desde então diversas aplicações foram desenvolvidas utilizando-se desta técnica, como análise de glóbulos brancos e cromossomos, análise automática de imagens de raios X, análise de artérias coronárias, classificação de impressões digitais, entre outras [23].

Na prática, o processo de esqueletização recebe como entrada uma imagem binária e produz uma imagem formada exclusivamente por linhas de um *pixel* de largura restando de maneira geral a configuração e conectividade da imagem original. O algoritmo utilizado foi o proposto por Zhang et al. [24] com as modificações propostas em [23]. A Figura 3.8 ilustra o processo de afinamento.

A segunda ferramenta que apresentamos para reconhecimento de padrões é conhecida como *crossing number*, conforme definido por Rutovitz [25]. O *crossing number* de Rutovitz,  $X(r)$ , é definido pela Equação 3.8 a partir da imagem binária resultante do processo de esqueletização, aqui representada pela função  $f(r)$ , em que  $r = (i, j)$ . Nesta imagem, o valor  $f(r)$  é 1 para pontos pertencentes às linhas resultantes ou zero caso contrário. Para cada ponto  $(i, j)$  da função  $f$  denota-se cada ponto adjacente tomado na ordem anti-horária a partir do ponto  $(i + 1, j)$  como  $n_{ij}(1), \dots, n_{ij}(8)$ . Dessa maneira, definimos  $\gamma(k)$  como o valor de  $f$  no  $k$ -ésimo vizinho desta sorte de um dado ponto  $(i, j)$ , ou seja,  $n_{ij}(k)$ . Se  $n_{ij}(k)$  extrapolar os limites da imagem, definimos  $\gamma(k) = 0$ .

$$X(r) = \sum_{k=1}^8 |\gamma(k+1) - \gamma(k)| \quad (3.8)$$

É resultante deste processamento a matriz  $X$ , de tamanho igual à imagem original, na qual cada ponto  $X(i, j)$  corresponde ao *crossing number* do ponto  $(i, j)$  na imagem resultante do processo de afinamento. A partir da matriz  $X$  é possível identificar pontos de interesse que contém informações sobre a morfologia da imagem original. Por exemplo, pontos em que  $X(i, j) = 2$  há uma terminação de linha. Esta informação, em conjunto com sua posição em relação às zonas predefinidas da imagem, é posteriormente utilizada como uma característica morfológica da imagem que é analisada.

### 3.6 Método para avaliação dos classificadores utilizados

Dado que o objetivo principal deste trabalho é o projeto e implementação de um classificador, é necessário definir o método utilizado para medir a eficácia dos sistemas propostos, tanto para comparar diferentes propostas e identificar qual delas produz melhores resultados quanto para comparar este trabalho a outros já publicados.

#### 3.6.1 *Precision e recall*

As duas mais básicas e frequentes medidas para eficácia de sistemas classificadores são *precision* e *recall* [26]. Definimos que um dos grupos classificados será nosso grupo alvo e o outro, de controle. Dado que: (i)  $tp$  (*true positive*) representa o número de classificações do grupo alvo corretamente realizadas; (ii)  $fp$  (*false positive*) representa o número de classificações do grupo alvo realizadas em erro; e (iii)  $fn$  (*false negative*) representa o número de classificação do grupo de controle realizadas em erro. Podemos definir *precision* e *recall* conforme as Equações 3.9 e 3.10.

$$\text{precision} = \frac{tp}{tp + fp} \quad (3.9)$$

$$\text{recall} = \frac{tp}{tp + fn} \quad (3.10)$$

A medida *precision* representa a relação entre itens do grupo alvo que foram corretamente classificados ( $tp$ ) e o total de itens classificados ( $tp + fp$ ). A medida *recall* representa a relação entre itens do grupo alvo que foram corretamente classificados ( $tp$ ) e o total de itens do grupo alvo ( $tp + fn$ ).

Em outras palavras, *precision* traduz a probabilidade de uma classificação como grupo alvo estar correta, enquanto *recall* traduz a probabilidade de um elemento do grupo alvo ser classificado como tal. A Tabela 3.1 auxilia no entendimento dos termos  $tp$ ,  $fp$ ,  $fn$  e  $tn$  ilustrando seus significados pelo cruzamento de linhas e colunas. A primeira coluna indica a classificação recebida

pelo elemento, enquanto a primeira linha indica a qual grupo o elemento de fato pertence.

Tabela 3.1: Classificação de elementos para avaliação de eficácia.

<i>Classificação:</i>	<i>Elemento do grupo alvo</i>	<i>Elemento do grupo de controle</i>
<i>Grupo alvo</i>	<i>tp true positive</i>	<i>fp false positive</i>
<i>Grupo de controle</i>	<i>fn false negative</i>	<i>tn true negative</i>

### 3.6.2 *F measure*

A vantagem de se obter ambos *precision* e *recall* se dá pois muitas vezes um é mais importante que o outro. Por exemplo, em nossa aplicação que classifica imagens que contém nudez, podemos favorecer *precision* caso seja importante que todas as imagens encontradas contenham de fato nudez, enquanto buscaremos melhor *recall* caso não seja admitido que imagens que contém nudez não sejam classificadas como tal.

No entanto, podemos utilizar uma única medida que traduz ambas características por meio de uma média harmônica ponderada, a *F measure*. A *F measure* é definida conforme a Equação 3.11, na qual  $\alpha \in [0, 1]$  e, portanto,  $\beta^2 \alpha \in [0, \infty]$ .

$$F_{measure} = \frac{1}{\alpha \frac{1}{precision} + (1 - \alpha) \frac{1}{recall}} = \frac{(\beta^2 + 1)precision \times recall}{\beta^2 precision + recall} \quad \text{onde} \quad \beta^2 = \frac{1 - \alpha}{\alpha} \quad (3.11)$$

Esta equação nos permite enfatizar *precision* ou *recall* em detrimento do outro para melhor adaptar o sistema proposto a cada objetivo. Neste caso, valores de  $\beta < 1$  enfatizam *precision* enquanto  $\beta > 1$  enfatizam *recall*. Neste trabalho todos os sistemas serão avaliados por meio da *F measure* com  $\beta = 1$  e, portanto, utilizando a Equação simplificada 3.12.

$$F_{measure} = \frac{2precision \times recall}{precision + recall} \quad (3.12)$$



## Capítulo 4

# Estudo da distribuição de cores de pele

A fim de atingir o objetivo deste trabalho, detectar nudez em imagens de maneira automática, a primeira tarefa foi definida como segmentar cor de pele nas imagens analisadas. A premissa é simples: corpos nus se apresentarão como objetos de cor de pele.

Conforme discutido na Seção 2.4, cores são um fenômeno de percepção humana e uma imagem digital colorida pode ser representada por meio da utilização de diversos espaços de cores diferentes. Estes espaços possuem características diferentes por terem sido concebidos com premissas e objetivos igualmente distintos.

Com o objetivo de classificar *pixels* como cor de pele ou não e a existência de diversos espaços de cores equivalentes, foi realizado um estudo da distribuição de *pixels* de cor de pele e *pixels* de controle nos espaços de cor  $RGB$ ,  $nRGB$ ,  $nRG$ ,  $HS$  e  $C_bC_r$  com o intuito de verificar o espaço mais adequado para realizar a segmentação. A escolha dos espaços de cor foi baseada em trabalhos anteriores [27, 28, 29, 30, 31, 32, 33].

### 4.1 Base de testes

Para realização de testes da distribuição de cores de pele, foi utilizada base de imagens criada por Jones e Rehg [34] também para definição de um modelo de cor de pele. A base é composta exclusivamente por imagens disponíveis à época na internet selecionadas sem filtros, inclusive contendo imagens de nudez. Por conta de sua variedade e sua popularidade na comunidade como referência para estudos desta sorte, consideramos a base ideal para os fins deste trabalho.

A base é composta por 4.670 imagens contendo pele e 8.965 imagens que não contém pele. As imagens que contém pele foram manualmente segmentadas por meio da utilização de uma ferramenta que permite ao usuário segmentar regiões de pele controlando um algoritmo de pontos conectados. Foi feito esforço para excluir olhos, cabelo e a abertura da boca na segmentação. O resultado é uma máscara binária que identifica os *pixels* de pele.

As imagens foram armazenadas nos formatos *GIF* ou *JPEG*. Por conta da degradação causada pela representação de poucos níveis de cor, desconsideramos as imagens *GIF* neste estudo. Após

desconsideradas as imagens inadequadas, trabalhamos com um total de 75.488.651 *pixels* de cor de pele e 653.800.137 de *pixels* de controle.

## 4.2 Distribuição dos grupos nos espaços de cor estudados

Nesta seção serão discutidos os resultados visuais da distribuição dos grupos de pele e controle nos espaços de cor  $RGB$ ,  $nRGB$ ,  $nRG$ ,  $HS$  e  $C_bC_r$ .

### 4.2.1 Distribuição no espaço de cor $RGB$

O espaço de cores  $RGB$ , apesar de ser muito utilizado para representação de imagens, possui a luminância distribuída entre seus três canais (*red*, *green* e *blue*). Justamente por isso possui baixa robustez a variações de luminosidade. Em nossa aplicação, tal característica é indesejável, uma vez que desejamos segmentar pele independentemente da luminosidade da amostra.

A fragilidade do espaço  $RGB$  foi melhor demonstrada por Saleh *et al.* em estudo que utilizou amostras de pele dos mesmos indivíduos com apenas variações nas condições de iluminação [27]. Conforme esperado, houve grande variação na distribuição nos canais  $RGB$ .

A Figura 4.1 representa a distribuição nos espaços  $RG$ ,  $RB$  e  $GB$  dos *pixels* das bases de pele e de controle. Nota-se os *pixels* de pele são distribuídos por ampla área dos espaços estudados, fenômeno que dificulta sua classificação por haver grande sobreposição com a distribuição do grupo de controle. No entanto, fica evidente que a distribuição dos *pixels* de pele é mais concentrada quando comparada ao grupo de controle, o que denota diferença entre os grupos propostos e possibilidade de classificação com a característica escolhida.

### 4.2.2 Distribuição nos espaços de cor $nRGB$ e $nRG$

O espaço de cor  $nRGB$ , conforme discutido no Capítulo 2, consiste em uma normalização direta do espaço de cores  $RGB$ . O objetivo da normalização é, entre outros, aumentar a robustez do espaço a variações de luminância.

A Figura 4.2 representa a distribuição nos espaços  $nRG$ ,  $nRB$  e  $nGB$  dos *pixels* das bases estudadas. Podemos notar os mesmos problemas citados para a distribuição no espaço  $RGB$ , como a distribuição do grupo de pele por grande área do espaço estudado e grande interseção com o grupo de controle.

No entanto, o espaço de cores  $nRG$  apresenta grande vantagem em relação ao  $RGB$  e ao  $nRGB$ , principalmente por conta de ser um espaço normalizado, com uma dimensão a menos e que não apresenta perda de informações (o componente descartado,  $b$ , é utilizado para normalização).

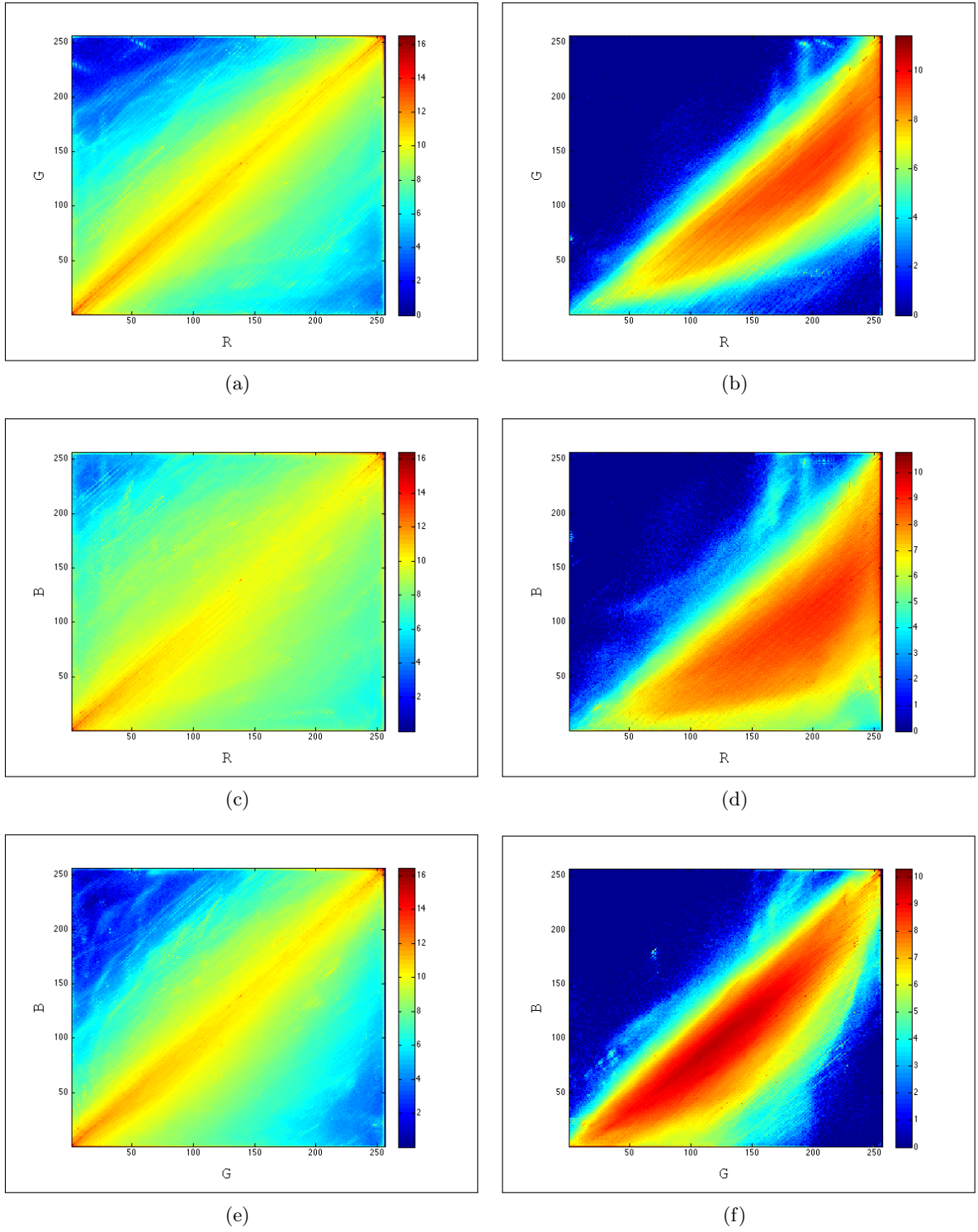
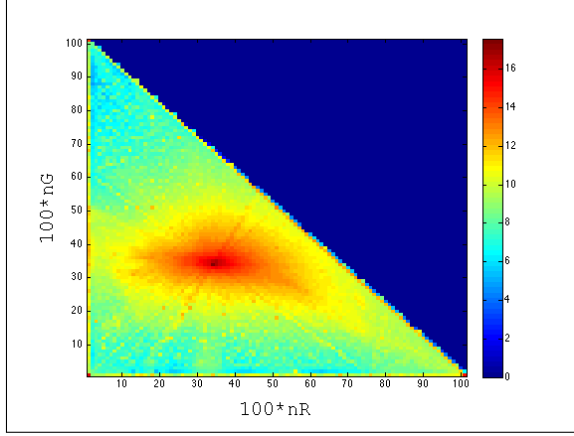
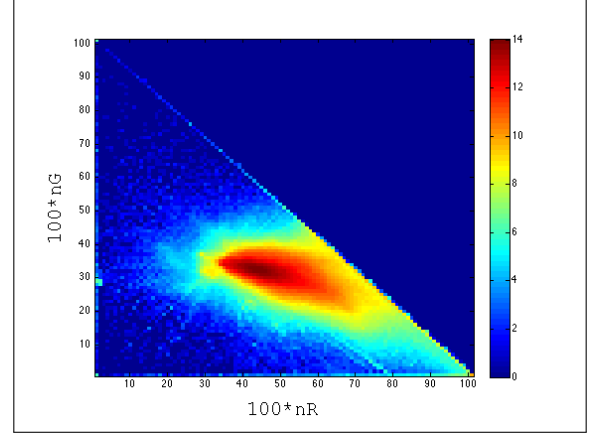


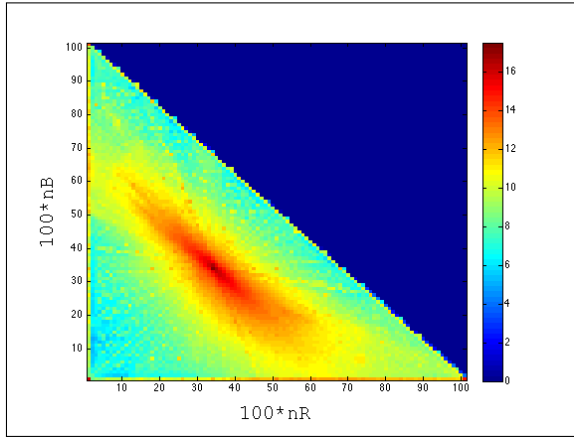
Figura 4.1: Distribuição de *pixels* de pele e não pele no espaço de cores  $RGB$ . (a), (c) e (e) representam a distribuição dos *pixels* que não são pele nos espaços  $RG$ ,  $RB$  e  $GB$ , respectivamente. (b), (d) e (f) representam a distribuição de *pixels* de pele nos espaços  $RG$ ,  $RB$  e  $GB$ , respectivamente.



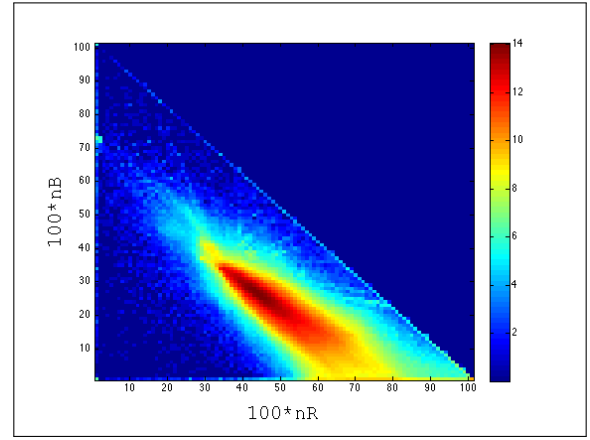
(a)



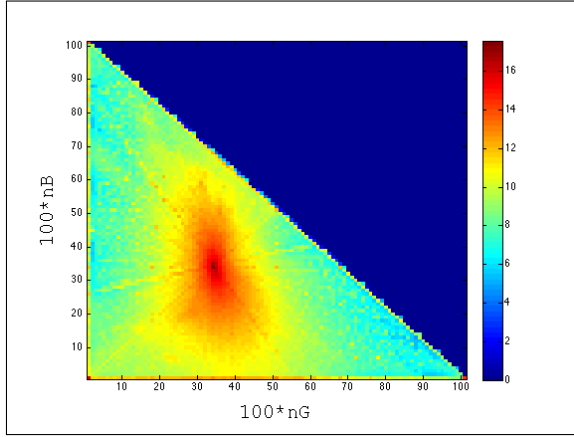
(b)



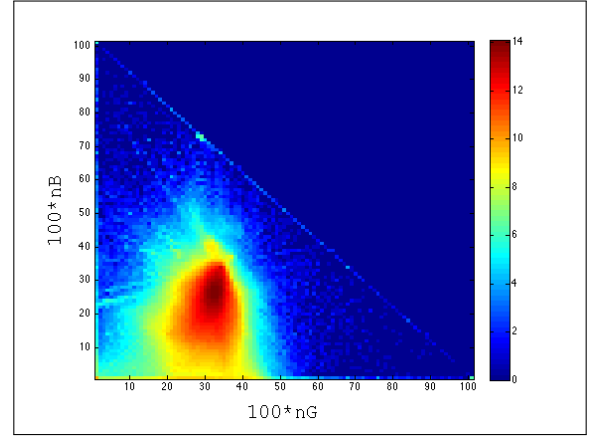
(c)



(d)



(e)



(f)

Figura 4.2: Distribuição de *pixels* de pele e não pele no espaço de cores *nRGB*. (a), (c) e (e) representam a distribuição dos *pixels* que não são pele nos espaços *nRG*, *nRB* e *nGB*, respectivamente. (b), (d) e (f) representam a distribuição de *pixels* de pele nos espaços *nRG*, *nRB* e *nGB*, respectivamente.

### 4.2.3 Distribuição no espaço de cor $HS$

O espaço de cor  $HS$  é formado pelos componentes *Hue* e *Saturation* do espaço  $HSV$ . Conforme discutido no Capítulo 2, esta representação é baseada em uma percepção humana das cores, procurando traduzir proximidade visual em proximidade no espaço.

A Figura 4.3 representa a distribuição no espaço de cor  $HS$  dos grupos de pele e de controle. É fácil notar que a distribuição de *pixels* de pele está concentrada nas duas extremidades horizontais do gráfico. Quando comparado aos espaços estudados até agora, tal característica já apresenta melhoria, por constituir distribuição mais diferenciada de seu respectivo grupo de controle, traduzindo-se diretamente em um problema mais simples de reconhecimento de padrões.

Trabalhos anteriores obtiveram bons resultados desconsiderando por completo o *cluster* menor de *pixels* de pele na distribuição  $HS$ , na Figura 4.3b à direita, considerando pele apenas os contidos no *cluster* maior, na mesma figura à esquerda. A dificuldade do espaço  $HS$  consiste em sua não-linearidade: conforme ilustrado na Figura 2.6, o componente *Hue* representa o ângulo do cilindro  $HSV$  causando, inclusive, a descontinuidade entre os dois *clusters* de pele identificados.

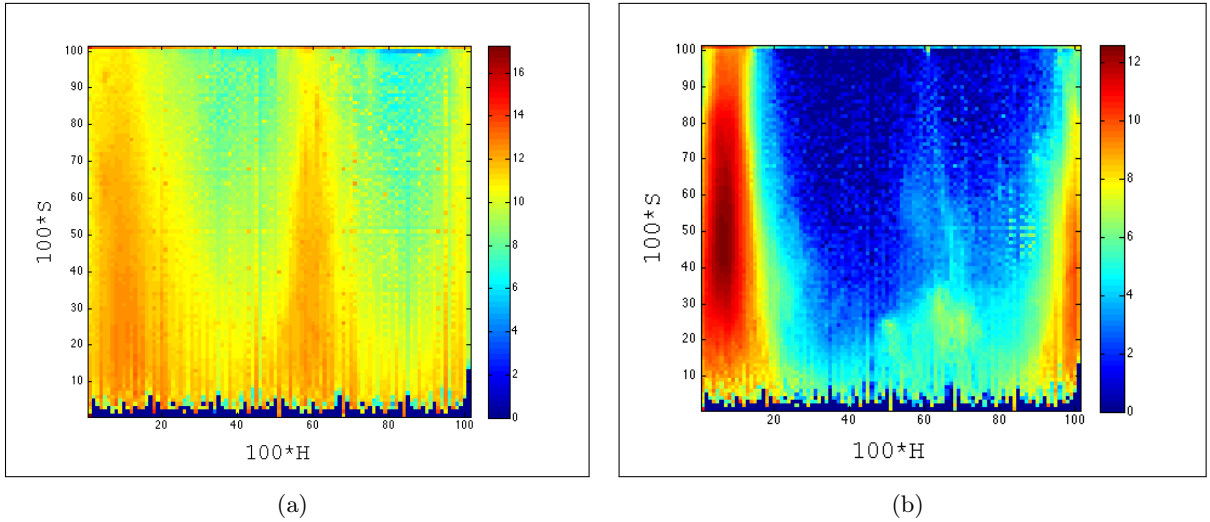


Figura 4.3: Distribuição de *pixels* de pele e não pele no espaço de cores  $HS$ . (a) representa a distribuição de *pixels* de controle enquanto (b) representa a distribuição de *pixels* de pele.

### 4.2.4 Distribuição no espaço de cor $C_bC_r$

O espaço  $YC_bC_r$  provê separação entre a luminância e a cromaticidade. De acordo com sua especificação, a luminância é isolada em seu componente  $Y$ , enquanto a cromaticidade é completamente representada no plano  $C_bC_r$ . Portanto, como estamos justamente buscando a independência da luminância, desconsideramos o componente  $Y$  e tratamos apenas o plano  $C_bC_r$ .

A Figura 4.4 representa a distribuição de *pixels* de pele e de controle no espaço  $C_bC_r$ . Da mesma forma que em outros espaços, os *pixels* de pele apresentam-se distribuídos em uma área mais restrita quando comparada aos de controle. No entanto, ainda há sobreposição entre eles.

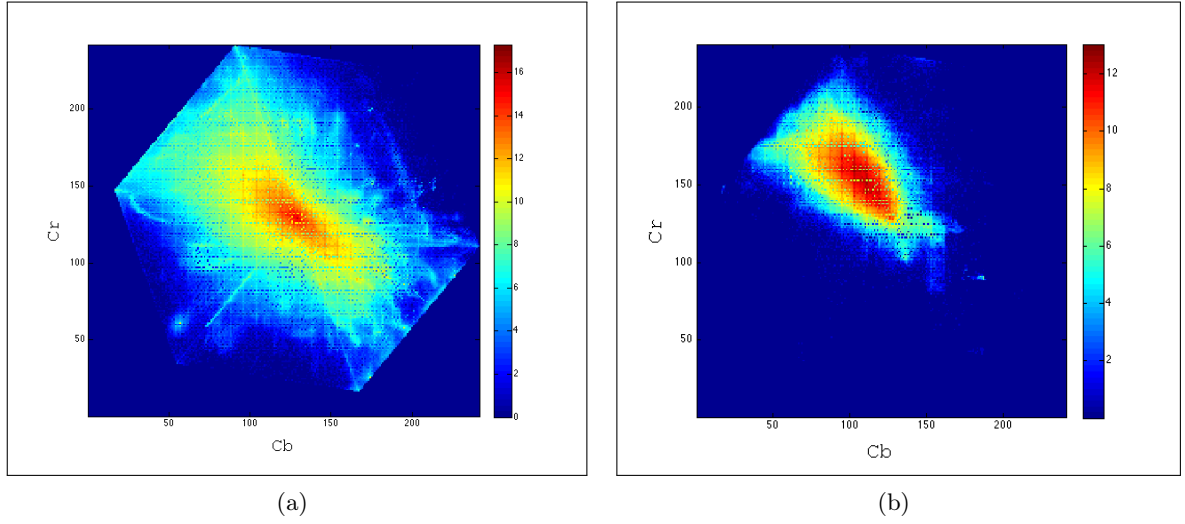


Figura 4.4: Distribuição de *pixels* de pele e não pele no espaço de cores  $C_bC_r$ . (a) representa a distribuição de *pixels* de controle enquanto (b) representa a distribuição de *pixels* de pele.

### 4.3 Segmentação de cor de pele por funções discriminantes lineares

Conforme discutido na Subseção 3.2.1, podemos utilizar funções discriminantes lineares para implementar um classificador que estabelece a base de comparação de performance para o novo classificador proposto. Esta Seção apresenta proposta de classificador deste tipo e os resultados obtidos.

#### 4.3.1 Classificador proposto

O classificador proposto utiliza a função discriminante linear equidistante dos centróides das distribuições de pele e de controle para classificar *pixels* em pele e não-pele. No caso de distribuições bidimensionais, como  $nRG$ ,  $C_bC_r$  e  $HS$ , o hiperplano de decisão do classificador é a reta dos pontos equidistantes dos centróides. No caso de distribuições tridimensionais como  $RGB$  e  $nRGB$ , o hiperplano de decisão é o plano dos pontos equidistantes de ambos os centróides.

A partir das distribuições calculadas e descritas na Seção 4.2 anterior, foram determinados os centróides dos *clusters* de pele e de controle a partir da média dos valores de cada *pixel* em cada dimensão do espaço estudado. Por exemplo, para o espaço de cores  $RGB$ , o centróide  $R_c, G_c, B_c$  tem seus componentes calculados por meio da média aritmética dos componentes de todos os  $n$  pontos, conforme descrito na Equação 4.1. O método é o mesmo para todos os espaços de cor estudados, calculando-se a média dos valores de cada um de seus componentes.

$$R_c = \sum_{k=1}^n \frac{R_k}{n}, \quad G_c = \sum_{k=1}^n \frac{G_k}{n}, \quad B_c = \sum_{k=1}^n \frac{B_k}{n} \quad (4.1)$$

Para realizar a classificação, os hiperplanos de decisão dividem o espaço em dois subespaços (pele e não-pele). Os pontos presentes no subespaço que contém o centróide da distribuição de

pontos de pele são classificados como pele. Os pontos presentes no subespaço que contém o centróide da distribuição de controle são classificados como não pele. Para efeitos desta classificação, consideramos os pontos contidos no hiperplano de decisão como não pele.

### 4.3.2 Resultados obtidos

Os resultados obtidos pela segmentação de cor de pele por menor distância aos centróides de pele e de controle foram insatisfatórios. No entanto, podemos extrair algumas informações úteis dos resultados, representados na Tabela 4.1. Os centróides de distribuição de pele e de controle são muito próximos em todas as distribuições, o que é traduzido em dificuldade de classificação. Além disso, ainda que muito aquém de qualquer classificação útil, nota-se melhor resultado com a utilização do espaço de cor  $C_bC_r$ , o que pode significar melhores condições gerais para classificação.

Tabela 4.1: Tabela representativa dos resultados obtidos com segmentação de pele por menor distância ao centróide.

<i>Espaço de Cor</i>	<i>Centróide C</i>	$C_{pele}$	$C_{no-pele}$	<i>F measure</i>
<i>RGB</i>	$\{R, G, B\}$	$\{184.29, 127.86, 98.19\}$	$\{117.53, 113.93, 107.12\}$	0.3175
<i>nRGB</i>	$\{nR, nG, nB\}$	$\{47.55, 31.67, 23.78\}$	$\{36.44, 34.08, 31.52\}$	0.4161
<i>nRG</i>	$\{nR, nG\}$	$\{47.55, 31.67\}$	$\{36.44, 34.08\}$	0.4387
<i>HS</i>	$\{H, S\}$	$\{11.64, 49.59\}$	$\{35.73, 36.09\}$	0.3788
$C_bC_r$	$\{C_b, C_r\}$	$\{106.63, 154.92\}$	$\{125.49, 131.07\}$	0.5301

## 4.4 Segmentação de cor de pele com redes neurais artificiais

Podemos facilmente notar, a partir dos resultados obtidos na Seção 4.3 e do estudo da distribuição de *pixels* de pele e controle conduzido na Seção 4.2, que este problema de classificação não possui solução linear com erro aceitável. Notamos inclusive que há grande interseção entre os pontos de pele e de controle em todas as distribuições. Este fenômeno faz sentido pois de fato a cor de pele humana faz parte das cores da natureza, representadas pelo grupo de controle.

Sabendo que a tarefa de classificação de *pixels* de pele não é linearmente separável ou sequer solucionável com baixo erro por conta da grande interseção entre estes e *pixels* do grupo de controle, buscamos a melhor alternativa de classificação por meio de redes neurais artificiais.

### 4.4.1 Classificador proposto

O classificador proposto é uma rede neural artificial (RNA) do tipo *feedforward* treinada utilizando-se o algoritmo Levenberg-Marquardt. Foi proposta uma arquitetura de três camadas, sendo uma de entrada, uma oculta e uma de saída. Na camada de entrada os neurônios não realizam processamento. Já nas camadas de saída e oculta é utilizada a função de ativação

tangente hiperbólica. Detalhes sobre o funcionamento da RNA com as características descritas são discutidos na Seção 3.3.

Na fase de treinamento, utilizou-se dois conjuntos de igual tamanho para treinamento e validação, com o objetivo de prevenir superaprendizado. Variou-se a quantidade de neurônios na camada escondida entre 4 e 19, de três em três. Com cada uma das quantidades de neurônios a rede foi inicializada com pesos aleatórios e treinada três vezes. Apenas a melhor rede entre todas as treinadas foi escolhida a partir da *F measure* calculada sobre sua classificação dos conjuntos de treinamento e validação. A fim de determinar a real eficácia da rede escolhida, foi calculada a *F measure* sobre um conjunto de testes que nunca havia sido apresentado à rede.

Na camada de saída foi utilizado apenas um neurônio. A função tangente hiperbólica utilizada nos neurônios das camadas oculta e de saída produz valores no intervalo  $[-1, 1]$ , de maneira que ela foi treinada com a saída 1 para os *pixels* de pele e  $-1$  para os de controle. A fim de se obter decisão binária deste classificador, a saída foi limiarizada, de maneira que valores no intervalo  $[0, 1]$  foram classificados como pele e valores no intervalo  $[-1, 0]$  como não pele.

#### 4.4.2 Resultados obtidos

A base de testes total disponível é composta por 75.488.651 *pixels* de cor de pele e 653.800.137 de *pixels* de controle. Por conta da grande diferença entre a quantidade de amostras de cada grupo e da significativa interseção entre os dois grupos, o primeiro passo foi subamostrar o grupo de *pixels* de controle selecionando 75.488.651 deles de forma aleatória.

Em seguida, ambos conjuntos foram divididos em três de igual tamanho: treinamento, validação e testes. A fim de minimizar o custo computacional, os conjuntos de treinamento e validação foram novamente subamostrados à taxa 500 : 1 resultando em 50.325 *pixels* de cada grupo em cada conjunto. A RNA foi então treinada conforme procedimento descrito na Subseção 4.4.1 anterior, com um máximo de 50 épocas de treinamento em cada tentativa.

Em seguida, a RNA treinada foi utilizada para classificar o conjunto de testes. Este não foi subamostrado, contendo 50.325.767 de *pixels* que nunca haviam sido apresentados à rede, sendo metade de pele e metade de controle. A partir da classificação pela RNA deste grupo, foi computada a *F measure*. A Tabela 4.2 apresenta os resultados obtidos pela melhor RNA em cada espaço de cor ao classificar o conjunto de testes.

Tabela 4.2: Análise de eficácia das RNAs de classificação de pele para diferentes espaços de cor

<i>Espaço de Cor</i>	<i>Neurônios</i>	<i>Épocas</i>	<i>Precision</i>	<i>Recall</i>	<i>F measure</i>
<i>RGB</i>	19	50	0.8545	0.9079	0.8804
<i>nRGB</i>	13	50	0.8254	0.9019	0.8620
<i>nRG</i>	7	50	0.8205	0.9047	0.8617
<i>HS</i>	13	50	0.8246	0.9050	0.8629
<i>C<sub>b</sub>C<sub>r</sub></i>	19	40	0.8545	0.9079	0.8804



Considerando a grande interseção entre os dois grupos que buscamos classificar, os resultados obtidos foram bem satisfatórios, com melhor *F measure* calculada para o espaço  $nRGB$ , de 0.8809. Ainda assim, os espaços  $RGB$  e  $C_bC_r$  apresentaram o mesmo resultado e muito próximo ao obtido, 0.8804. Os demais espaços de cor também apresentaram bons resultados, porém inferiores. Vale ressaltar que o espaço  $C_bC_r$  foi o único bidimensional com resultado praticamente equivalente aos espaços tridimensionais  $RGB$  e  $nRGB$ , confirmando a hipótese estabelecida anteriormente de ser uma boa alternativa para segmentação de cores de pele.

Por conta da proximidade entre os resultados obtidos, decidimos por prosseguir com os testes utilizando todos os espaços de cor, com o objetivo de buscar o melhor resultado possível prevenindo descartes precipitados.

## Capítulo 5

# Método proposto

Este capítulo descreve o método proposto para *classificação automática de nudez em imagens*. A Figura 5.1 apresenta um diagrama ilustrativo da solução proposta, em alto nível de abstração. O sistema desenvolvido possui como entrada apenas uma imagem digital colorida no padrão *RGB*, sem metadados. A entrada é submetida ao processamento que pode ser dividido em três subsistemas: (i) Classificador de cor de pele; (ii) Extrator de características; e (iii) Classificador de nudez.

Em primeiro lugar, é realizada comparação com a solução proposta em trabalho previamente publicado [35]. Em seguida, são descritos em detalhes cada um dos subsistemas que compõe esta solução de classificação de imagens.

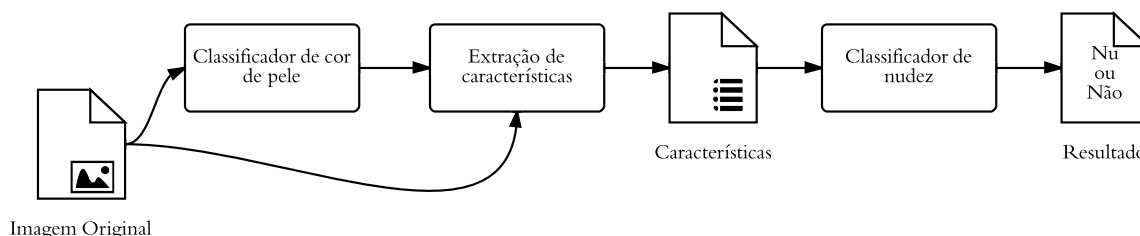


Figura 5.1: Diagrama ilustrativo do método proposto.

### 5.1 Solução proposta anteriormente

Em [35], foi proposta solução de detecção de nudez baseada em redes neurais artificiais. Sua composição pode também ser ilustrada pela Figura 5.1, exceto pela extração de características ser realizada apenas a partir da saída do classificador de pele, sem utilizar a imagem original. Nos próximos parágrafos serão brevemente descritas as diferenças entre a solução proposta anteriormente e a proposta neste trabalho.

O classificador de cor de pele do sistema anterior foi desenvolvido com base na observação empírica da distribuição de 226.188 *pixels* de pele sobre o espaço de cor *HS*. Este classificador determinava pele *pixels* como pele no caso  $H \in [0, 0.11]$  e  $S \in [0.2, 0.7]$  e não-pele caso contrário. Em seguida, as mesmas operações morfológicas de abertura e fechamento foram empregadas.

Por outro lado, o classificador de cor de pele aqui proposto é baseado em redes neurais artificiais treinadas a partir da base de dados Jones-Rehg [34] com mais de 75 milhões de *pixels* de pele e 653 *pixels* de controle. Foram utilizados os espaços de cor *RGB*, *nRGB*, *nRG* e *C<sub>b</sub>C<sub>r</sub>* além do *HS* já utilizado anteriormente. Mais informações sobre o classificador de pele proposto neste trabalho constam no Capítulo 4 e na Seção 5.2.

Em seguida, a extração de características proposta anteriormente utilizou apenas duas: (i) número de elementos conectados de *pixels* de pele; e (ii) área relativa do maior elemento conectado. O classificador aqui proposto utiliza cinco principais características e propõe melhorias nas utilizadas anteriormente: (i) número de elementos conectados de *pixels* de pele; (ii) área relativa dos dois maiores elementos conectados; (iii) área relativa de pele em zonas; (iv) área relativa acumulada de pele em faces; e (v) características morfológicas dos grupos de pele.

Finalmente, o classificador de nudez aqui proposto também é baseado em redes neurais artificiais. No entanto, além de utilizar maior número de características, foi treinado com o dobro de amostras quando comparado ao anterior. Ainda mais, a fim de determinar as melhores características para classificação, foi utilizado rigoroso protocolo de testes, este descrito no Capítulo 6.

## 5.2 Classificador de cor de pele

A primeira etapa do método proposto baseia-se na segmentação por cor de pele da imagem apresentada. A entrada deste classificador é a imagem original. A saída é uma matriz bidimensional em que cada ponto  $(x, y)$  possui o valor *zero* caso o *pixel* da imagem original na posição  $(x, y)$  não tenha sido classificado como pele ou o valor indicativo do elemento conectado de *pixels* de pele ao qual ele pertence caso contrário. Nesta Seção este processamento é descrito em detalhes.

Com o objetivo de redução da complexidade computacional e, por conseguinte, o tempo de processamento das imagens, é realizado redimensionamento da imagem original. A imagem é redimensionada mantendo-se a proporção entre suas dimensões originais de maneira que a maior delas, altura ou largura, passe a medir 300 *pixels*. Esta operação de pré-processamento permite que imagens de alta resolução ( $5 + MP$ ) possam ser analisadas pelo sistema em tempo razoável sem prejuízo de eficácia do classificador [6]. A imagem com dimensões reduzidas é a utilizada em todos os passos subsequentes.

Para segmentação de grupos de pele, o primeiro passo é a classificação, *pixel a pixel*, de cor de pele. Para este fim, foram utilizados os classificadores descritos em detalhes no Capítulo 4, baseados em redes neurais artificiais supervisionadas do tipo *feedforward* com Levenberg–Marquardt. Dado um vetor de entrada que contém os valores de cada componente de cor de um *pixel*, a rede o classifica como pele ou não. Foram treinadas diversas redes para cada um dos espaços de cor

utilizados e delas escolhida a com melhor eficácia, esta medida pela *F measure*. Foram utilizados os espaços de cor *RGB*, *nRGB*, *nRG*, *C<sub>b</sub>C<sub>r</sub>* e *HS*.

Portanto, na etapa de classificação, a imagem *RGB* é primeiro convertida para o espaço de cor que será utilizado, conforme métodos descritos na Seção 2.4. Em seguida, cada *pixel* da imagem é apresentado à rede neural em termos de suas componentes de cor. O resultado da rede é inserido em uma máscara binária, com o valor de *zero* para indicar *pixels* de pele e *um* caso contrário. A Figura 5.2b ilustra esta máscara obtida pela classificação da imagem original Figura 5.2a.

Em seguida, são realizadas as operações morfológicas de abertura seguida de fechamento pelo elemento estruturante círculo de raio 5 *pixels*. Esta etapa tem por objetivo eliminar *pixels* isolados ou grupos muito pequenos que tenham sido classificados e unir grupos de pele muito próximos. A Figura 5.2c ilustra o resultado dessas operações morfológicas.

Uma vez concluídas a classificação e as operações morfológicas, é obtida uma máscara que indica *pixels* de pele com o valor *zero* e com *um* os demais *pixels*. A partir desta máscara é necessário rotular cada elemento conectado de *pixels* de pele a fim de possibilitar, por exemplo, contar o número de elementos ou determinar a área de cada um. É resultado da etapa de rotulação uma matriz de rótulos na qual o valor de cada elemento indica a qual elemento conectado ele pertence, utilizando-se *zero* para *pixels* de não-pele. O Algoritmo 5.1 é utilizado para rotular os elementos conectados. Vale notar que, em uma segunda passagem, estabelece-se a equivalência entre rótulos, quando regiões vizinhas possuem rótulos diferentes.

---

**Algoritmo 5.1** Rotulamento de elementos conectados.

---

```

1  R(i, j) = 0;
2  foreach M(i, j) == 0 do
3      if M(i, j-1) == 1 && M(i-1, j) == 1
4          indice = indice + 1;
5          R(i, j) = indice;
6      elseif M(i, j-1) == 0 && M(i-1, j) == 1
7          R(i, j) = R(i, j-1);
8      else
9          R(i, j) = R(i, j-1);
10     end
11 end

```

---

A Figura 5.3 ilustra o processo de classificação de pele por completo, a partir da imagem original representada na Figura 5.3a. O resultado da classificação de *pixels* de pele após as operações morfológicas é representado pela Figura 5.3b. A partir deste resultado são rotulados os elementos conectados de pele, representados pelos diferentes tons de cinza e branco na Figura 5.3c. Nesta figura os *pixels* que não foram classificados como pele estão em preto.

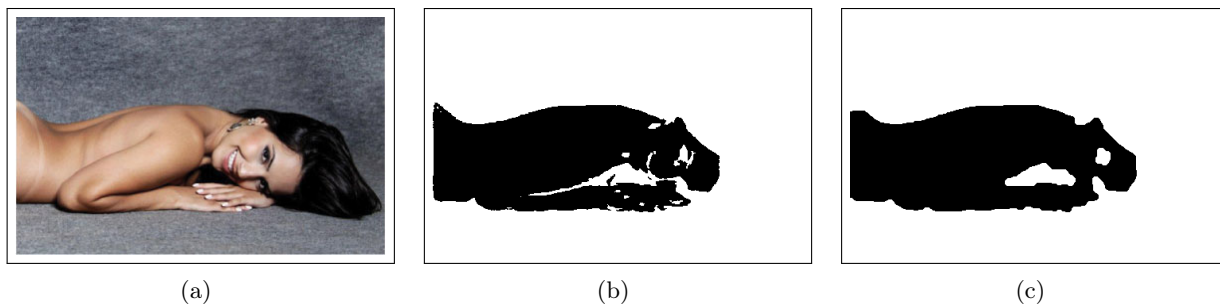


Figura 5.2: Exemplo da segmentação por cor de pele. (a) representa a imagem original submetida ao processamento; (b) representa o resultado da classificação de *pixels* de pele, estando estes indicados em preto; e (c) é o resultado final da segmentação após realização de operações morfológicas.

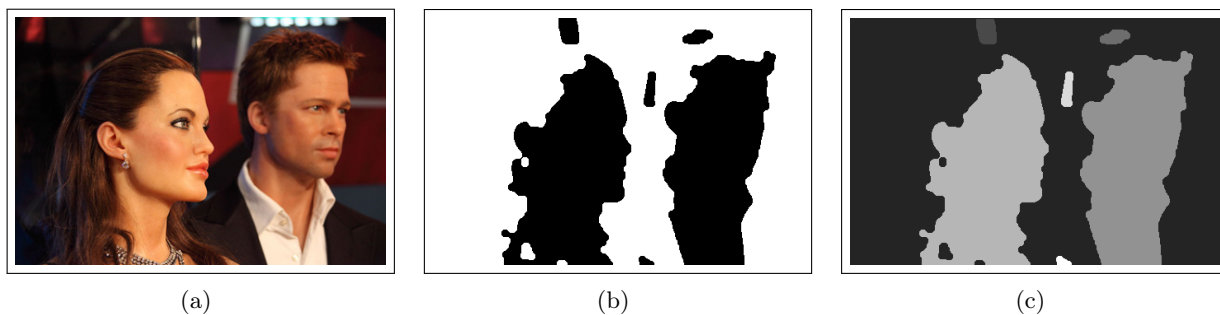


Figura 5.3: Exemplo de identificação de elementos conectados. (a) representa a imagem original submetida ao processamento; (b) representa o resultado da classificação de *pixels* de pele após operações morfológicas (c) ilustra os grupos de *pixels* de pele conectados em tons de cinza diferentes e *pixels* de não-pele em cor preta.

Figura 5.3a em domínio público, disponível em [publicdomainpictures.net](http://publicdomainpictures.net).

## 5.3 Extrator de características

Conforme discutido no Capítulo 3, uma das tarefas mais importantes do reconhecimento de padrões é a escolha das características das amostras que serão utilizadas para que se determine a qual classe pertencem. O padrão que se deseja identificar é o de imagens que contém nudez. Devemos, portanto, identificar características das imagens que indicam a presença de nudez. Foram extraídas características a partir da imagem original, da máscara de *pixels* de pele e da matriz de elementos conectados.

Nesta Seção serão exploradas, uma a uma, as características identificadas e utilizadas para tal.

### 5.3.1 Número de elementos conectados de área significativa

A primeira característica analisada é o número e a área dos elementos conectados de *pixels* de pele presentes na imagem. A hipótese estabelecida é que imagens que contenham nudez possuirão menos corpos de pele porém de maior área, enquanto imagens com muitas faces ou com segmentação incorreta de *pixels* de pele apresentarão maior número de corpos de menor área.

Esta característica é diretamente extraída da saída do classificador de pele, com os elementos conectados identificados e rotulados, conforme ilustrado pela Figura 5.3c. A partir da matriz de elementos conectados  $R$ , é calculado o vetor de áreas relativas destes elementos, `groupAreas`. Este vetor possui o comprimento igual ao número de elementos conectados e cada posição armazena a área de um dos elementos, em ordem decrescente. O Algoritmo 5.2 descreve a função que o calcula.

---

**Algoritmo 5.2** Medição de área de elementos conectados.

---

```
1 function calculateGroupAreas(R)
2     [h, w] = size(R);
3     pixelCount = h*w;
4     groupCount = max(R);
5     groupAreas = zeros(groupCount);
6     for i = 1:groupCount
7         groupAreas(i) = count(R == i)/pixelCount;
8     end
9     return revert(sort(groupAreas));
10 end
```

---

A partir do vetor de áreas `groupAreas` podemos calcular quantos elementos conectados possuem área superior a determinado limiar  $\rho$  por meio da expressão `count(groupAreas >  $\rho$ )`. A partir de testes discutidos em detalhes na Seção 6.3.1, determinou-se que o valor  $\rho = 0$  é o que logra melhores resultados de classificação e, portanto, utilizamos a contagem de todos os elementos conectados como característica, independente de sua área.

### 5.3.2 Área relativa dos dois maiores elementos conectados

Esta característica objetiva extrair informações sobre o tamanho que elementos conectados de pele ocupam na imagem. A hipótese avaliada é que, ao selecionar os maiores elementos conectados, haveria grande diferença entre casos de nudez, com elementos de grande área e não-nudez, com elementos menores e mais fragmentados. A área relativa foi utilizada para que não houvesse variação com imagens de diferentes dimensões.

A área relativa de cada elemento conectado é calculada pelo Algoritmo 5.2, que retorna um vetor em ordem decrescente da área de cada elemento. A partir dele são extraídas duas características: (i) área relativa do maior elemento conectado; e (ii) área relativa do segundo maior elemento conectado. A determinação do número  $n$  de maiores elementos dos quais a área foi utilizada como característica foi determinada de forma empírica, conforme descrito na Seção 6.3.2. Tais experimentos determinaram que os melhores resultados são obtidos com  $n = 2$ , ou seja, com a área dos dois maiores elementos conectados.

### 5.3.3 Área relativa de pele em zonas

Geralmente em uma imagem a região central é a mais representativa enquanto que regiões mais externas representam cenários ou outros elementos menos importantes [36]. A técnica de zoneamento já foi utilizada inclusive para classificadores de nudez por cor de pele [37].

Partindo desta hipótese, a imagem foi seccionada em três zonas cujas fronteiras foram determinadas por dois retângulos concêntricos de dimensões  $\frac{4}{9}h \times \frac{4}{9}w$  e  $\frac{2}{3}h \times \frac{2}{3}w$ , sendo que  $h$  e  $w$  representam a altura e a largura da imagem, respectivamente. A Zona  $Z_3$ , mais central, é delimitada pela área interna ao primeiro retângulo. A Zona  $Z_2$ , intermediária, é delimitada pela área entre o segundo retângulo e o primeiro. A Zona  $Z_1$ , mais externa, é delimitada pela área entre a fronteira da imagem e o segundo retângulo. A Figura 5.4 ilustra este processo de zoneamento da imagem.

Com as zonas estabelecidas, calcula-se a área relativa  $A_{Z_n}$  de *pixels* de pele em cada zona  $Z_n$  por meio da Equação 5.1.

$$A_{Z_n} = \frac{\text{número de } \textit{pixels} \text{ de pele em } Z_n}{\text{número de } \textit{pixels} \text{ em } Z_n} \quad (5.1)$$

Foi determinado empiricamente que os melhores resultados de classificação são obtidos com a área relativa ocupada por *pixels* de pele em cada uma das três zonas, conforme discutido na Seção 6.3.3. Portanto, utilizamos como características  $A_{Z_1}$ ,  $A_{Z_2}$  e  $A_{Z_3}$ .

### 5.3.4 Características de faces presentes na imagem

A presença de faces humanas e cor de pele em imagens estão fortemente relacionadas. Diversos trabalhos publicados utilizam classificação de *pixels* de pele para implementar ou melhorar detectores de faces [34, 38, 39, 40]. Portanto, foi levantada a hipótese de que a presença de faces poderia

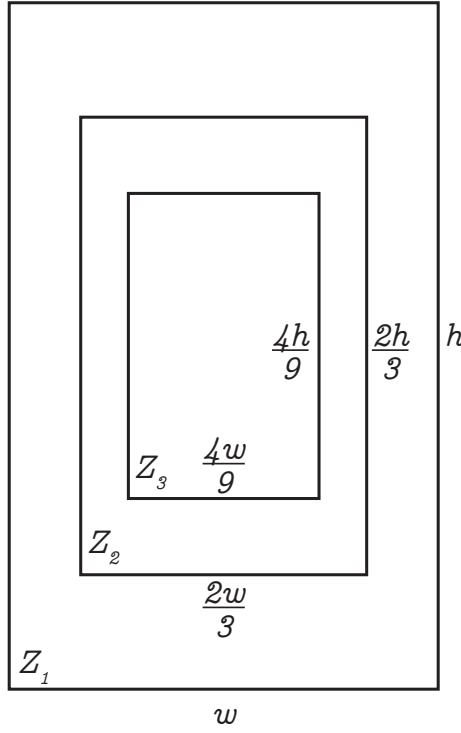


Figura 5.4: Definição de zonas da imagem.

causar grande número de falsos positivos na classificação. Com o objetivo de confirmar esta hipótese e mitigar falsos positivos gerados por faces, foram projetadas três características extraídas da saída do classificador de faces em conjunto com a máscara de *pixels* de pele.

O classificador de faces utilizado foi o clássico Viola-Jones, melhor descrito na Seção 3.4. Este possui como entrada a imagem original e como saída um vetor de retângulos que delimitam as faces encontradas. Cada retângulo é indicado com sua posição na imagem  $(x, y)$ , sua largura  $w$  e sua altura  $h$ . A partir deste vetor e da máscara de *pixels* de pele extraem-se as características utilizadas.

A primeira, mais simples, nada mais é que o *número de faces detectadas na imagem*. Esta baseia-se na hipótese de que quanto mais faces presentes em uma imagem, menor a probabilidade de conter nudez, visto que a grande maioria destas imagens possui entre zero e duas faces. Seu cálculo consiste na contagem de elementos do vetor saída do classificador de faces.

A segunda consiste na *área relativa acumulada de faces*. Sua utilização é baseada na hipótese de que imagens cuja proporção de área ocupada por faces é muito elevada (*e.g.* retratos) não contém nudez. O cálculo da área relativa acumulada de faces  $Af$ , é realizado por meio da Equação 5.2, onde  $k$  é o número de elementos do vetor de saída do classificador de faces;  $h_i$  e  $w_i$  são, respectivamente, a altura e largura do retângulo delimitador da face detectada  $i$ ; e  $h$  e  $w$  são, respectivamente, a altura e largura da imagem.

$$Af = \sum_{i=1}^k \frac{h_i \times w_i}{h \times w} \quad (5.2)$$



Por fim, a terceira característica consiste na *área relativa acumulada de pele em faces*. Esta consiste na relação entre o número de *pixels* de pele dentro de retângulos delimitadores das faces encontradas e o número total de *pixels* de pele na imagem. Seu cálculo é realizado por meio do Algoritmo 5.3, no qual  $M$  representa a máscara de *pixels* de pele e  $faces$  representa o vetor de saída do classificador de faces. A Figura 5.5 ilustra o processamento realizado para extração desta característica que, neste caso, pode ser entendida como a razão entre o número de *pixels* azuis e o número de *pixels* azuis e pretos.

---

**Algoritmo 5.3** Cálculo da área de pele em faces.

---

```

1  function calculateSkinAreaInFaces(M, faces)
2      [h, w] = size(M);
3      facesCount = length(faces);
4      skinPixels = sum(sum(M));
5      skinInFaces = 0.0;
6      for i = 1:facesCount
7          skinInFaces += sum(sum(iscrop(M, faces(i))))/skinPixels;
8      end
9      return skinInFaces;
10 end

```

---

Conforme descrito na Seção 6.3.4, foi determinado empiricamente que a utilização apenas da terceira característica, *área relativa acumulada de pele em faces*, fornece os melhores resultados de classificação, permitindo-nos descartar as outras duas projetadas.

### 5.3.5 Formas de elementos conectados de pele

Por fim, a última característica projetada tem como objetivo extrair informações da morfologia assumida pelos corpos de pele presentes na imagem. Este processo tem como entrada a máscara de *pixels* de pele e como saída o número de elementos morfológicos presentes na imagem.

O primeiro passo consiste no afinamento da máscara de pele por meio do processo descrito na Seção 3.5. O resultado é uma imagem binária formada exclusivamente por linhas de 1 *pixel* de espessura que preservam a configuração original dos corpos de pele detectados.

Em seguida, calcula-se a matriz de *cross number* de Rutovitz, conforme descrito na Seção 3.5. Esta matriz é então utilizada para identificação de pontos de interesse em relação à morfologia dos corpos: pontos em que há bifurcações ou terminações de linhas.

Com o objetivo de preservar alguma informação sobre a posição dos elementos morfológicos de interesse foi utilizado o método de zoneamento descrito na Seção 5.3.3, contando-se separadamente os elementos encontrados em cada zona. O número de elementos morfológicos encontrados em cada uma das zonas  $Z_1$ ,  $Z_2$  e  $Z_3$  constituem as três características de formas de elementos conectados de pele. A Figura 5.6b ilustra o resultado deste processamento na imagem da Figura 5.6a, de maneira que as linhas resultantes do processamento morfológico estão na cor preta, os elementos

morfológicos encontrados estão assinalados com quadrados vermelhos e os retângulos delimitadores das zonas estão em verde.

Foi determinado de forma empírica que os melhores resultados são obtidos quando considerada apenas a contagem de elementos morfológicos na Zona  $Z_3$ , conforme descrito na Seção 6.3.5.

## 5.4 Classificador de nudez

Por fim, o classificador de nudez proposto consiste em uma rede neural artificial do tipo *feed-forward* treinada de forma supervisionada por meio do algoritmo Levenberg–Marquardt, conforme detalhado na Seção 3.3. Os neurônios utilizam a função de transferência tangente hiperbólica e são organizados em três camadas. A primeira não realiza processamento, apenas responsável por receber as entradas e encaminhar adiante. A segunda, camada oculta, tem entre 4 e 19 neurônios. A terceira, camada de saída, possui apenas um neurônio.

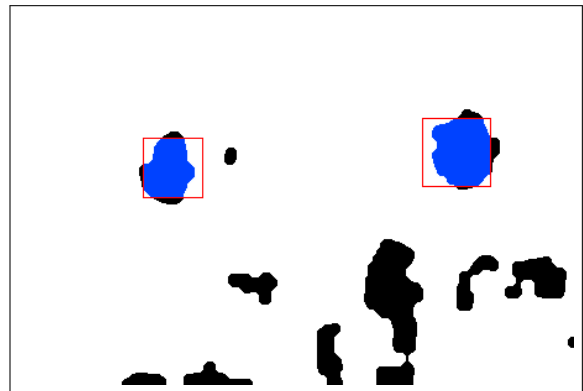
A RNA recebe como parâmetro um vetor das características que lograram melhores resultados nos testes descritos no Capítulo 6 e fornece como saída um valor entre  $-1$  e  $1$ . A classificação é realizada pela limiarização da saída desta RNA, classificando a imagem como de nudez caso a saída seja maior que *zero* ou como não nudez caso contrário.

O vetor de características utilizado para classificação foi determinado empiricamente, conforme descrito no Capítulo 6 a seguir. Este vetor é composto pelas características: (i) número de elementos conectados de *pixels* de pele; (ii) área relativa do maior elemento conectado; (iii) área relativa do segundo maior elemento conectado; (iv) porcentagem de pele na zona  $Z_1$ ; (v) porcentagem de pele na zona  $Z_2$ ; (vi) porcentagem de pele na zona  $Z_3$ ; (vii) porcentagem do total de pele que está presente em faces detectadas; (viii) quantidade de elementos morfológicos presentes na zona  $Z_3$ .

O método proposto está representado por completo no diagrama da Figura 5.7, com detalhamento de cada característica utilizada para classificação e qual processamento necessário para obtê-la.



(a)

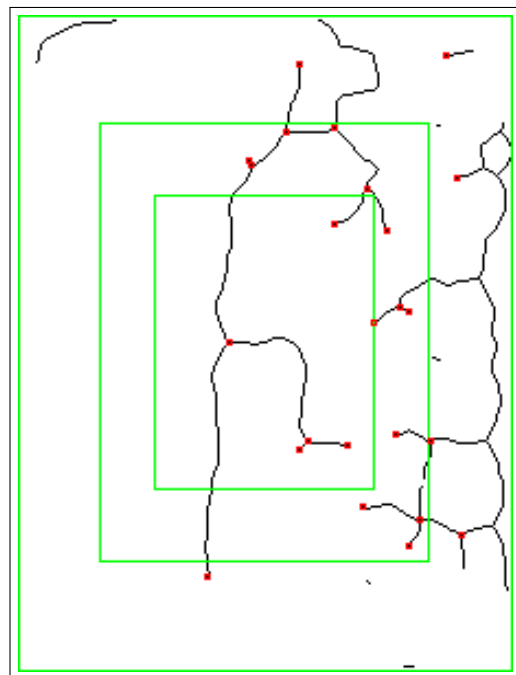


(b)

Figura 5.5: Exemplo de extração da característica de *pixels* de pele presentes em faces. (a) representa a imagem de entrada com as faces detectadas assinaladas; (b) representa o processamento realizado sobre a máscara de pele, sendo os *pixels* de pele dentro de faces assinalados em azul e os demais *pixels* de pele assinalados em preto.



(a)



(b)

Figura 5.6: Exemplificação do processo de extração de características morfológicas.

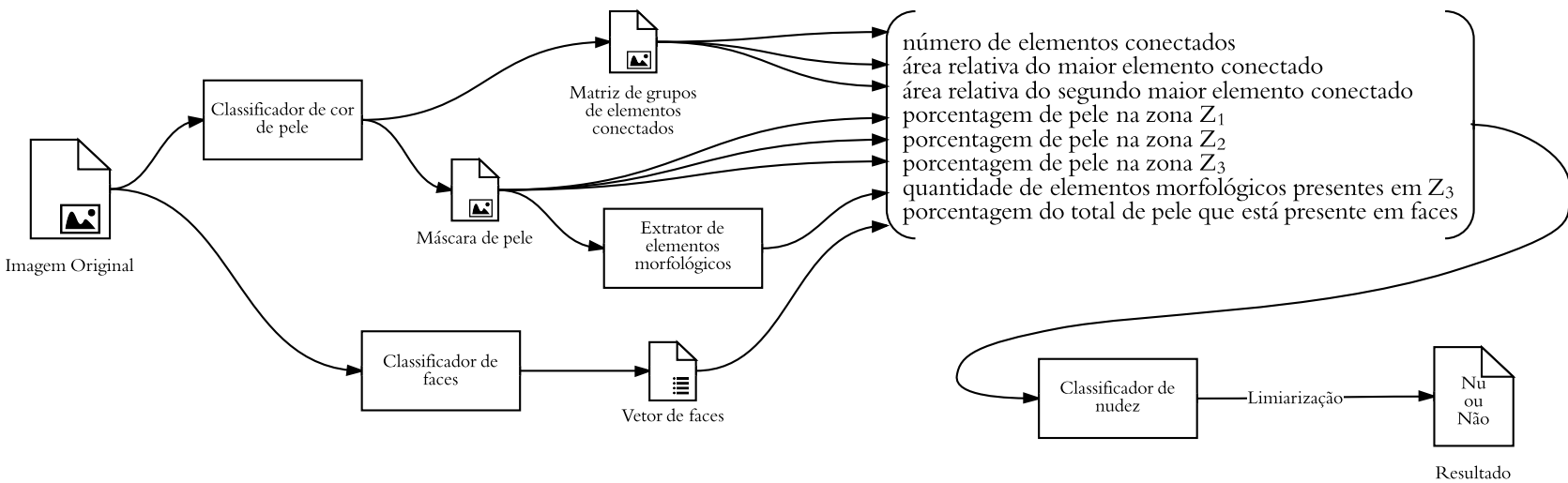


Figura 5.7: Método proposto completo com detalhamento das características utilizadas.

## Capítulo 6

# Resultados Obtidos

Este capítulo apresenta os testes projetados para avaliar a eficácia de cada componente do sistema proposto e do sistema como um todo, bem como os resultados obtidos em cada teste. Também é apresentado o conjunto de testes utilizado.

### 6.1 Conjunto de testes

O conjunto de testes utilizado possui 1.000 imagens que contém nudez e 1.000 imagens que não contém. As imagens pornográficas foram obtidas por buscas na internet, em arquivos de *torrent* e fóruns. O principal critério utilizado na coleta foi a presença de partes íntimas expostas e o aparência amadora das imagens (evitamos coletar imagens profissionais). As imagens não pornográficas consistem em uma miscelânea de fotos particulares, fotos disponíveis em redes sociais e imagens sintetizadas. O critério de seleção utilizado buscou manter número muito maior de fotos reais e primar por fotos em que pessoas apareciam, como retratos ou fotos de grupos. Também foram incluídas fotos que contém muitos falsos positivos para cor de pele, como praias e madeiras.

Este conjunto, com o dobro de amostras, contém todas as utilizadas em [35]. Como o classificador proposto é baseado em redes neurais artificiais de aprendizado supervisionado, este conjunto foi dividido em três subconjuntos, treinamento, validação e teste, cada um com um terço das amostras de ambos os grupos. Vale notar que a ordem dos elementos foi aleatorizada antes de cada teste, a fim de se evitar viés ao criar os subconjuntos.

### 6.2 Protocolo de testes

Os testes realizados visam determinar o melhor conjunto de características e parâmetros do classificador de nudez proposto. Para tanto, subdividimos os testes em duas fases: *(i)* avaliação de características; e *(ii)* avaliação do sistema proposto.

A fase de *avaliação de características* tem como objetivo determinar o melhor conjunto de características das descritas na Seção 5.3. Para tal, definimos um classificador de nudez básico,

com apenas duas características, o número de elementos conectados de pele e a área relativa do maior elemento conectado. A partir deste classificador básico, cada característica foi adicionada ou modificada e a eficácia do sistema aferida. Por conta do grande número de características é inviável a avaliação de todas suas possíveis combinações e, portanto, cada nova característica foi avaliada isoladamente em termos de seu impacto sobre o classificador básico.

Determinado o melhor conjunto de características, a fase de *avaliação do sistema proposto* objetiva treinar a rede neural mais eficaz a partir da variação do segmentador de pele utilizado, entre os espaços de cor  $RGB$ ,  $nRGB$ ,  $nRG$ ,  $C_bC_r$  e  $HS$ . Esta fase determina o sistema completo que obtém melhor eficácia de classificação.

Com o objetivo de se obter o melhor classificador possível, em cada teste a rede neural classificadora de nudez foi treinada com 4, 7, 10, 13, 16 e 19 neurônios na camada escondida, sendo inicializada e treinada três vezes para cada quantidade de neurônios. A rede com melhor *F measure* foi selecionada.

## 6.3 Fase $i$ – Avaliação de características

Esta Seção descreve a fase  $i$  de testes, que objetiva determinar, entre todas as características propostas no Capítulo 5 anterior, qual a combinação que resulta no classificador de nudez em imagens mais eficaz.

### 6.3.1 Área mínima para contagem de elementos conectados

Este teste tem como objetivo definir o limiar  $\rho$  ótimo, ou seja, a área relativa mínima considerada significativa para extração da característica do *número de elementos conectados de área significativa*, conforme descrito na Seção 5.3.1.

A RNA utilizada neste teste possui como vetor de entrada apenas duas características. A primeira consiste na área relativa do maior elemento conectado. A segunda é justamente o número de elementos conectados de área significativa, variando-se o limiar  $\rho$ . Os resultados obtidos estão representados na Tabela 6.1, com o melhor classificador destacado.

Determinamos, portanto, que a limiarização da área significativa para contagem de elementos conectados de pele não contribui com a classificação, tendo sido o melhor resultado obtido quando considerados todos os elementos conectados. Este resultado estabeleceu o classificador de referência para os testes de análise de características a seguir, com a *F measure* base 0.8296.

### 6.3.2 Área relativa dos maiores elementos conectados

Este teste tem como objetivo avaliar o impacto da utilização das áreas relativas dos segundo e terceiro maiores elementos conectados no classificador proposto. o primeiro teste consistiu em adicionar ao vetor de características do classificador base ( $n = 1$ ) a área do segundo maior elemento

conectado ( $n = 2$ ). O segundo teste consistiu em adicionar as áreas de ambos o segundo e o terceiro maiores elementos conectados ( $n = 3$ ).

Os resultados obtidos estão representados na Tabela 6.2. Conforme esperado, adicionar a área relativa do segundo maior elemento aumentou significativamente a eficácia do classificador. No entanto, ao adicionar também a área do terceiro maior elemento, a eficácia piorou inclusive comparada ao classificador base. Consideramos, portanto, a utilização apenas da área dos dois maiores elementos conectados como características para o classificador final ( $n = 2$ ).

### 6.3.3 Área relativa de pele em zonas da imagem

Este teste tem como objetivo determinar o impacto das características baseadas na área relativa de zonas da imagem ocupadas por *pixels* de pele, conforme descrito na Seção 5.3.3. A fim de escolher o conjunto ótimo de características, todas as combinações foram exploradas e sua eficácia comparada ao classificador base. Os resultados obtidos estão representados na Tabela 6.3.

Conforme esperado, a utilização desta característica aumentou a eficácia do classificador, obtendo *F measure* de 0.8416, comparado a 0.8296 obtido pelo classificador base. No entanto, a melhora foi observada apenas quando utilizadas as áreas relativas de peles de todas as três zonas –  $Z_1$ ,  $Z_2$  e  $Z_3$  – como características. Vale notar que a utilização das áreas de cada zona isoladamente ou qualquer combinação de duas zonas piorou a eficácia do classificador, um resultado importante por ser inesperado.

### 6.3.4 Características de faces presentes na imagem

Este teste tem como objetivo avaliar o impacto na eficácia do classificador quando utilizadas as características extraídas a partir dos detectores de faces e de pele discutidas na Seção 5.3.4: (i) número de faces presentes na imagem; (ii) quantidade de *pixels* de pele em faces em relação ao total de *pixels* de pele; e (iii) área relativa da imagem ocupada por faces.

A fim de determinar a melhor combinação entre estas características, todas foram avaliadas e os resultados compilados na Tabela 6.4. Conforme esperado, adicionar características relacionadas às faces presentes na imagem apresentou grande melhoria na eficácia do classificador. No entanto, a utilização de apenas a característica ii – *quantidade relativa de pixels de pele em faces* – apresentou resultado superior, com *F measure* de 0.8555, quando comparado às demais combinações. As demais características – i e iii – foram, portanto, descartadas.

### 6.3.5 Características da forma dos elementos conectados

Este teste tem como objetivo avaliar o impacto na classificação de nudez quando utilizadas as características da contagem de elementos morfológicos descrita na Seção 5.3.5. Esta contagem foi realizada separadamente para as zonas  $Z_1$ ,  $Z_2$  e  $Z_3$  definidas na Seção 5.3.3. A fim de determinar a melhor combinação de características, todas as combinações foram avaliadas. Os resultados obtidos estão representados na Tabela 6.5. Também foi avaliada a hipótese de não se utilizar o zoneamento

Tabela 6.1: Análise de eficácia com diferentes limiares  $\rho$  para área significativa.

$\rho$	<i>Neurônios</i>	<i>Precision</i>	<i>Recall</i>	<i>F measure</i>
0%	13	0.8359	0.8234	0.8296
5%	13	0.7934	0.7934	0.7934
10%	7	0.7507	0.8293	0.7881
15%	13	0.7784	0.8413	0.8086
20%	10	0.7910	0.8383	0.8140
25%	4	0.7976	0.7904	0.7940

Tabela 6.2: Análise de eficácia com área dos  $n$  maiores elementos conectados.

$n$	<i>Neurônios</i>	<i>Precision</i>	<i>Recall</i>	<i>F measure</i>
1	13	0.8359	0.8234	0.8296
2	19	0.8362	0.8713	0.8534
3	7	0.8313	0.7964	0.8135

Tabela 6.3: Análise de eficácia com área de pele por em zonas.

<i>Zonas</i>	<i>Neurônios</i>	<i>Precision</i>	<i>Recall</i>	<i>F measure</i>
—	13	0.8359	0.8234	0.8296
$Z_1$	19	0.8278	0.8204	0.8241
$Z_2$	19	0.8006	0.8174	0.8089
$Z_3$	16	0.8313	0.8263	0.8288
$Z_1, Z_2$	19	0.8355	0.7754	0.8043
$Z_1, Z_3$	10	0.8277	0.8054	0.8164
$Z_2, Z_3$	16	0.8179	0.8204	0.8191
$Z_1, Z_2, Z_3$	13	0.8742	0.8114	0.8416

Tabela 6.4: Análise de eficácia com características de faces na imagem.

<i>Características</i>	<i>Neurônios</i>	<i>Precision</i>	<i>Recall</i>	<i>F measure</i>
—	13	0.8359	0.8234	0.8296
$i$	4	0.8558	0.8353	0.8455
$ii$	4	0.8348	0.8772	0.8555
$iii$	13	0.8174	0.8713	0.8435
$i, ii$	4	0.8056	0.8683	0.8357
$i, iii$	16	0.8358	0.8533	0.8444
$ii, iii$	16	0.8092	0.8383	0.8235
$i, ii, iii$	10	0.8478	0.8503	0.8490



da imagem, tomando como característica apenas a contagem total de elementos morfológicos. Esta alternativa está representada na Tabela 6.5 pela alternativa “ $Z_1 + Z_2 + Z_3$ .”

Foi possível obter melhoria na classificação com a utilização da contagem de elementos morfológicos. No entanto, esta melhoria só foi verificada quando considerada apenas a contagem de elementos morfológicos presentes na área central da imagem, ou seja, na zona  $Z_3$ . Foram, portanto, descartados os elementos morfológicos externos a  $Z_3$ .

### 6.3.6 Conclusão da fase $i$ de testes

A fase  $i$  de testes teve como objetivo determinar quais características comporiam o vetor utilizado como entrada do último componente do sistema proposto: a rede neural classificadora de nudez. Este objetivo foi cumprido a partir da observação de resultados por vezes já antecipados e outras vezes surpreendentes.

O vetor de características definido por meio deste processo contém oito características no total, sendo essas:  $(i)$  número de elementos conectados de *pixels* de pele;  $(ii)$  área relativa do maior elemento conectado;  $(iii)$  área relativa do segundo maior elemento conectado;  $(iv)$  porcentagem de pele na zona  $Z_1$ ;  $(v)$  porcentagem de pele na zona  $Z_2$ ;  $(vi)$  porcentagem de pele na zona  $Z_3$ ;  $(vii)$  porcentagem da área total de pele que está presente em faces detectadas;  $(viii)$  quantidade de elementos morfológicos presentes na zona  $Z_3$ .

## 6.4 Fase $ii$ – Avaliação do sistema proposto

Esta seção descreve a fase  $ii$  de testes, que objetiva determinar a melhor rede neural artificial para classificação de nudez a partir das características definidas pela fase  $i$  de testes. Com o sistema proposto estruturado e o vetor de características já estabelecido, realizaram-se os testes substituindo apenas o detector de pele utilizado, a fim de determinar qual o espaço de cores mais adequado para classificação, entre  $RGB$ ,  $nRGB$ ,  $nRG$ ,  $HS$  e  $C_bC_r$ .

Os resultados obtidos para o sistema proposto estão representados na Tabela 6.6. Apesar do estudo realizado no Capítulo 4 apontar todos os espaços de cores analisados como bons candidatos para segmentação de pele, quando analisado o sistema proposto por completo a eficácia dos outros espaços de cor são muito inferiores à do  $RGB$ .

O classificador proposto, portanto, obteve  $F$  *measure* de 0.8735.

Tabela 6.5: Análise de eficácia com informações da forma dos elementos conectados.

<i>Zonas</i>	<i>Neurônios</i>	<i>Precision</i>	<i>Recall</i>	<i>F measure</i>
—	13	0.8359	0.8234	0.8296
$Z_1 + Z_2 + Z_3$	13	0.8018	0.8114	0.8065
$Z_1$	19	0.8281	0.7934	0.8104
$Z_2$	10	0.7902	0.8234	0.8065
$Z_3$	4	0.8383	0.8383	0.8383
$Z_1, Z_2$	19	0.8043	0.7754	0.7896
$Z_1, Z_3$	16	0.8428	0.8024	0.8221
$Z_2, Z_3$	19	0.8138	0.8114	0.8126
$Z_1, Z_2, Z_3$	4	0.8360	0.7934	0.8141

Tabela 6.6: Eficácia da solução proposta com diferentes espaços de cor para classificação de nudez.

<i>Espaço de Cor</i>	<i>Neurônios</i>	<i>Precision</i>	<i>Recall</i>	<i>F measure</i>
<i>RGB</i>	16	0.9013	0.8473	0.8735
<i>nRGB</i>	10	0.8379	0.8204	0.8290
<i>nRG</i>	10	0.8529	0.8503	0.8516
<i>HS</i>	16	0.8701	0.8623	0.8662
<i>C<sub>b</sub>C<sub>r</sub></i>	19	0.7972	0.8473	0.8215

## Capítulo 7

# Conclusões

### 7.1 Objetivo geral

Neste trabalho foi apresentado um método para classificação de imagens que contenham nudez baseado em redes neurais artificiais. Por meio de dois classificadores baseados em RNAs, um para *pixels* de cor de pele e outro para nudez, obtiveram-se resultados satisfatórios para classificação de nudez.

O método proposto obteve *F measure* de 0.8735, classificando 90.1% das imagens com nudez corretamente e com 84.7% de chance de uma imagem classificada de fato conter nudez. Este resultado é muito satisfatório e permitiria o uso deste sistema em ambiente real, seja para filtros de conteúdo ou para investigações policiais, conforme sugerido no Capítulo 1.

Vale ressaltar ainda que o sistema proposto utiliza apenas operações morfológicas simples e classificadores altamente eficientes (RNAs com única camada oculta e o detector de faces Viola-Jones). A implementação utilizada, codificada e executada em MATLAB, teve como objetivo a prova de conceito do método proposto e sua execução não é eficiente. Portanto, não foram apresentados testes de eficiência e capacidade de processamento. No entanto, com uma implementação eficiente a simplicidade do método permitiria sua utilização em sistemas de filtragem de conteúdo em tempo real ou para perícia em campo com recursos limitados.

A principal contribuição deste trabalho é um método eficiente baseado em redes neurais artificiais para classificação de imagens que contenham nudez. Consideramos cumprido, portanto, o principal objetivo deste trabalho.

### 7.2 Outras contribuições

Neste trabalho foi proposta e testada a utilização de um detector de faces para classificação de nudez, partindo da hipótese de que cor de pele e faces estão intimamente relacionadas e, portanto, a presença de faces pode gerar falsos positivos. Esta hipótese foi confirmada pelo aumento de eficácia do classificador quando utilizada a relação entre *pixels* de pele presentes em faces e os

*pixels* de pele presentes na imagem como característica.

Adicionalmente, foi proposta e testada característica baseada no que chamamos de elementos morfológicos calculados a partir do *crossing number* de Rutovitz após a esqueletização da máscara de pele (ver Seção 5.3.5). Este método é geralmente utilizado para identificação de minúcias em impressões digitais porém foi utilizado com sucesso para aprimorar a classificação de nudez quando aliado ao zoneamento da imagem.

Por fim, foi apresentado amplo estudo de classificação de pele em cinco espaços de cores diferentes:  $RGB$ ,  $nRGB$ ,  $nRG$ ,  $HS$  e  $C_bC_r$ . A partir desse estudo foi proposto um classificador consistente apenas de RNA e comparados os resultados para todos os espaços de cor estudados. Para tal utilizou-se a vasta base de imagens Jones-Rehg. Apesar da eficácia desses classificadores ter sido muito próxima isoladamente (variação máxima da  $F$  *measure* de 0.018, ver Tabela 4.2), verificou-se grande discrepância quando analisado o sistema por completo (ver Tabela 6.6). Determinamos, portanto, que os espaços de cor mais indicados para classificação de pele são o  $RGB$  e o  $HS$ .

## 7.3 Trabalhos futuros

O principal trabalho futuro deste projeto é implementar de maneira eficiente o método proposto a fim de possibilitar seu uso em aplicações reais, bem como medir a eficiência desta implementação. Esta tarefa pode ser realizada, por exemplo, pela utilização da ferramenta OpenCV<sup>1</sup>, um *toolkit* de operações em imagens altamente eficiente.

Também citamos como trabalho futuro a utilização de classificadores baseados em *deep learning* a partir das características identificadas para comparação com os resultados obtidos até agora, buscando melhorar a eficácia do método.

---

<sup>1</sup><http://opencv.org>

# REFERÊNCIAS BIBLIOGRÁFICAS

- [1] UCKO, P. J.; ROSENFELD, A. *Palaeolithic cave art*. [S.l.]: McGraw-Hill Nueva York, 1967.
- [2] GERNESHEIM, H. *A concise history of photography*. [S.l.]: Courier Dover Publications, 1986.
- [3] ISO/IEC 10918-1:1994 - Information technology – Digital compression and coding of continuous-tone still images: Requirements and guidelines.
- [4] FACEBOOK; ERICSSON; QUALCOMM. *A Focus on Efficiency*. [S.l.], 2013. Disponível em: [https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-prn1/851575\\_520797877991079\\_393255490\\_n.pdf](https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-prn1/851575_520797877991079_393255490_n.pdf).
- [5] HOUAISS, A. et al. *Dicionário Houaiss da língua portuguesa*. [S.l.]: Objetiva, 2009. ISBN 9788573029635.
- [6] POLASTRO, M. de C.; ELEUTERIO, P. da S. Nudetective: A forensic tool to help combat child pornography through automatic nudity detection. In: *Database and Expert Systems Applications (DEXA), 2010 Workshop on*. [S.l.: s.n.], 2010. p. 349–353. ISSN 1529-4188.
- [7] GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2002. ISBN 978-0201180756.
- [8] CAI, J.; GOSHTASBY, A.; YU, C. Detecting human faces in color images. In: *International Workshop on Multi-Media Database Management Systems, 1998. Proceedings*. [S.l.: s.n.], 1998. p. 124–131.
- [9] HUNKE, M.; WAIBEL, A. Face locating and tracking for human-computer interaction. In: *1994 Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers, 1994*. [S.l.: s.n.], 1994. v. 2, p. 1277–1281 vol.2.
- [10] TERRILLON, J.-C.; DAVID, M.; AKAMATSU, S. Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments. In: *Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings*. [S.l.: s.n.], 1998. p. 112–117.
- [11] IMAGAWA, K.; LU, S.; IGI, S. Color-based hands tracking system for sign language recognition. In: *Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings*. [S.l.: s.n.], 1998. p. 462–467.

- [12] XIONG, Y.; FANG, B.; QUEK, F. Extraction of hand gestures with adaptive skin color models and its applications to meeting analysis. In: *Eighth IEEE International Symposium on Multimedia, 2006. ISM'06*. [S.l.: s.n.], 2006. p. 647–651.
- [13] WYSZECKI, G.; STILES, W. S. *Color science*. [S.l.]: Wiley New York, 1982.
- [14] LANG, L.-y.; GU, W.-W. Study on face detection algorithm based on skin color segmentation and adaboost algorithm. In: *Web Mining and Web-based Application, 2009. WMWA'09. Second Pacific-Asia Conference on*. [S.l.]: IEEE, 2009. p. 70–73.
- [15] SKARBEEK, W. lady's law; KOSCHAN, A. Colour image segmentation a survey. 1994.
- [16] POYNTON, C. Frequently asked questions about color. *Retrieved June*, v. 19, p. 2004, 1997.
- [17] FLECK, M. M.; FORSYTH, D. A.; BREGLER, C. *Finding Naked People*. [S.l.: s.n.], 1996.
- [18] ROSSOTTI, H. Colour: Why the world isn't grey. Penguin Books Ltd., 1983.
- [19] DUDA, R.; HART, P.; STORK, D. *Pattern classification*. [S.l.]: Wiley, 2001. (Pattern Classification and Scene Analysis: Pattern Classification). ISBN 9780471056690.
- [20] MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943.
- [21] DEMUTH, H. et al. *Neural Network Toolbox For Use with Matlab*. 1993.
- [22] VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *IEEE. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. [S.l.], 2001. v. 1, p. I–511.
- [23] LAM, L.; LEE, S.-W.; SUEN, C. Thinning methodologies-a comprehensive survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 14, n. 9, p. 869–885, Sep 1992. ISSN 0162-8828.
- [24] ZHANG, T.; SUEN, C. Y. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, ACM, v. 27, n. 3, p. 236–239, 1984.
- [25] RUTOVITZ, D. Pattern recognition. *Journal of the Royal Statistical Society. Series A (General)*, Wiley for the Royal Statistical Society, v. 129, n. 4, p. pp. 504–530, 1966. ISSN 00359238. Disponível em: <<http://www.jstor.org/stable/2982255>>.
- [26] MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to information retrieval*. [S.l.]: Cambridge university press Cambridge, 2008.
- [27] SALEH, A.; SAGHEER, A. Automatic skin detection in color images. Disponível em: <<https://www.stlo.unicaen.fr/icisp/ct/uploads.old/882/1-ICISP2012.pdf>>.
- [28] YANG, J.; LU, W.; WAIBEL, A. Skin-color modeling and adaptation. In: . [S.l.: s.n.], 1997. p. 687–694.

- [29] ZARIT, B. D.; SUPER, B. J.; QUEK, F. K. Comparison of five color models in skin pixel classification. In: *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on.* IEEE, 1999. p. 58–63. Disponível em: <[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=799224](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=799224)>.
- [30] VEZHNEVETS, V.; SAZONOV, V.; ANDREEVA, A. A survey on pixel-based skin color detection techniques. In: *Proc. Graphicon.* Moscow, Russia, 2003. v. 3, p. 85–92. Disponível em: <<http://academic.aau.dk/Skhachat/Public/Papers%20on%20Face%20Detection/Survey%20on%20Skin%20Color%20Techniques.pdf>>.
- [31] LU, J. et al. A comparative study of skin-color models. In: *Image Analysis and Recognition.* [S.l.]: Springer, 2005. p. 729–736.
- [32] PHUNG, S. L.; BOUZERDOUM, A.; SR, D. C. Skin segmentation using color pixel classification: analysis and comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 27, n. 1, p. 148–154, 2005.
- [33] ELGAMMAL, A.; MUANG, C.; HU, D. Skin detection. In: *Encyclopedia of Biometrics.* [S.l.]: Springer, 2009. p. 1218–1224.
- [34] JONES, M. J.; REHG, J. M. Statistical color models with application to skin detection. *International Journal of Computer Vision*, Springer, v. 46, n. 1, p. 81–96, 2002.
- [35] ZAGHETTO, A.; SANDOVAL, D.; FRANCO, P. Detecção de nudez utilizando redes neurais artificiais. In: *ANAIS DA ICMEDIA.* [s.n.], 2012. p. 40–43. Disponível em: <<http://www.icmedia.org.br/2012/images/stories/arquivos/ANAIS-ICMEDIA2012.pdf>>.
- [36] KALVA, P.; ENEMBRECK, F.; KOERICH, A. Web image classification using classifier combination. *Latin America Transactions, IEEE (Revista IEEE America Latina)*, v. 6, n. 7, p. 661–671, Dec 2008. ISSN 1548-0992.
- [37] SANTOS, C.; SANTOS, E. dos; SOUTO, E. Nudity detection based on image zoning. In: *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on.* [S.l.: s.n.], 2012. p. 1098–1103.
- [38] SEOW, M.-J.; VALAPARLA, D.; ASARI, V. Neural network based skin color model for face detection. In: *Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. 32nd.* [S.l.: s.n.], 2003. p. 141–145.
- [39] WANG, X. et al. Robust real-time face detection with skin color detection and the modified census transform. In: *Information and Automation, 2008. ICIA 2008. International Conference on.* [S.l.: s.n.], 2008. p. 590–595.
- [40] SAGHEER, A.; ALY, S. An effective face detection algorithm based on skin color information. In: *Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on.* [S.l.: s.n.], 2012. p. 90–96.

# ANEXOS



# I. DESCRIÇÃO DO CONTEÚDO DO CD

O conteúdo do CD é organizado da maneira a seguir:

- `monografia.pdf` – este documento, texto completo da monografia em formato PDF;